UNIVERSITY OF SOUTHAMPTON

Identification by a Hybrid 3D/2D Gait Recognition Algorithm

by

Fatimah Shamsulddin Abdulsattar

A thesis submitted in partial fulfilment for the
degree of Doctor of Philosophy

in the
Faculty of Physical and Applied Sciences
Electronics and Computer Science

December 2016

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL AND APPLIED SCIENCES

Electronics and Computer Science

Doctor of Philosophy

IDENTIFICATION BY A HYBRID 3D/2D GAIT RECOGNITION ALGORITHM

by Fatimah Shamsulddin Abdulsattar

Recently, the research community has given much interest in gait as a biometric. However, one of the key challenges that affects gait recognition performance is its susceptibility to view variation. Much work has been done to deal with this problem. The implicit assumptions made by most of these studies are that the view variation in one gait cycle is small and that people walk only along straight trajectories. These are often wrong. Our strategy for view independence is to enrol people using their 3D volumetric data since a synthetic image can be generated and used to match a probe image.

A set of experiments was conducted to illustrate the potential of matching 3D volumetric data against gait images from single cameras inside the Biometric Tunnel at Southampton University using the Gait Energy Image as gait features. The results show an average Correct Classification Rate (CCR) of 97% for matching against affine cameras and 42% for matching against perspective cameras with large changes in appearance. We modified and expanded the Tunnel systems to improve the quality of the 3D reconstruction and to provide asynchronous gait images from two independent cameras. Two gait datasets have been collected; one with 17 people walking along a straight line and a second with 50 people walking along straight and curved trajectories.

The first dataset was analysed with an algorithm in which 3D volumes were aligned according to the starting position of the 2D gait cycle in 3D space and the sagittal plane of the walking people. When gait features were extracted from each frame using Generic Fourier Descriptors and compared using Dynamic Time Warping, a CCR of up to 98.8% was achieved. A full performance analysis was performed and camera calibration accuracy was shown to be the most import factor. The shortcomings of this algorithm were that it is not completely view-independent and it is affected by changes in walking directions.

A second algorithm was developed to overcome the previous limitations. In this, the alignment was based on three key frames at mid-stance phase. The motion in the first and second parts of the gait cycle was assumed to be linear. The second dataset was used for evaluating the algorithm and a CCR of 99% was achieved. However, when the probe consisted of people walking on a curved trajectory, the CCR dropped to 82%. But when the gallery was also taken from curved walking, the CCR returned to 99%. The algorithm was also evaluated using data from the Kyushu University 4D Gait Database where normal walking achieved 98% and curved walking achieved 68%. Inspection of the data indicated that the assumption made previously that straight ahead walking and curved walking are similar, is invalid. Finally, an investigation into more appropriate features was also carried out but this only gave a slight improvement.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, **Fatimah Shamsulddin Abdulsattar**, declare that the thesis entitled **Identification by a hybrid 3D/2D Gait Recognition Algorithm** and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research.

I confirm that:

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

- Fatimah Abdulsattar and John Carter. ***Performance Analysis of Gait Recognition with Large Perspective Distortion***. Accepted to be published in the IEEE International Conference on Identity, Security and Behaviour Analysis, Japan, 2016.
- Fatimah Abdulsattar and John Carter. ***A Practical Technique for Gait Recognition on Curved and Straight Trajectories***. Accepted to be published in the 9th IAPR International Conference on Biometrics, Sweden, 2016.

Signed: ____

Date: ____

Acknowledgements

I would like to express my deep gratitude to my supervisor Dr. John Carter for his valuable advice and guidance throughout my PhD thesis. His support and encouragement gave me a light of hope when the path appeared to be dark and hard. Without his assistance and time, modifying and expanding the Tunnel system have been an extremely difficult task.

I am grateful thank to the PhD students of Electronics and Computer Science School at Southampton University for their participation in collecting different data which was not only been used in the analysis but also in characterising several issues in the old Tunnel system.

My deep thanks also go to my husband and parents for their support and encouragement throughout my study.

Finally, I would like to thank the Higher Committee for Education Development in Iraq for their financial support of my scholarship.

To my husband, parents, sisters, brother and lovely kids.

Chapter 1

Introduction

Human identification and verification play an important role in many fields of modern societies, ranging from authorising financial transactions to fighting crime and illegal immigration. The traditional identification documents, such as passports, bank cards and driving licenses, provide a convenient and easy method for personal identification, but they are unreliable because they can be easy misplaced, forgotten or forged. In many high throughput areas where a high level of security and reliability is required, biometrics technology can be used to provide automatic verification of a person's identity based on fingerprints, iris patterns, DNA or facial expressions. Recently, gait has emerged as a promising biometric cue for human recognition because of its attractive characteristics particularly for surveillance scenarios. Gait recognition aims at identifying people according to their walking styles. Recognition can be performed based on the information extracted from the static shape and/or motion over a sequence of image frames.

Recently, recognising people by their gait has achieved a very high recognition rate under controlled conditions [42]. However, one of the common challenges of gait recognition is view variations as the appearance of an individual changes under different views and consequently, the recognition performance is adversely affected [101, 58]. Fig. 1.1 shows how the appearance of the individual changes under different views.

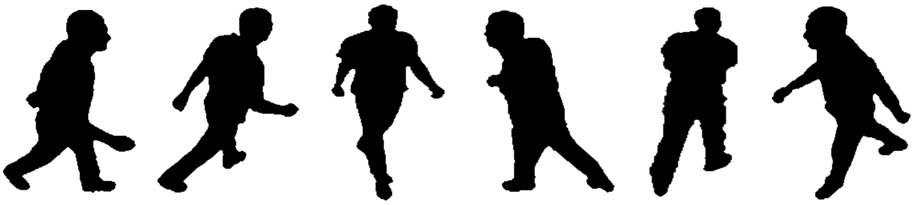


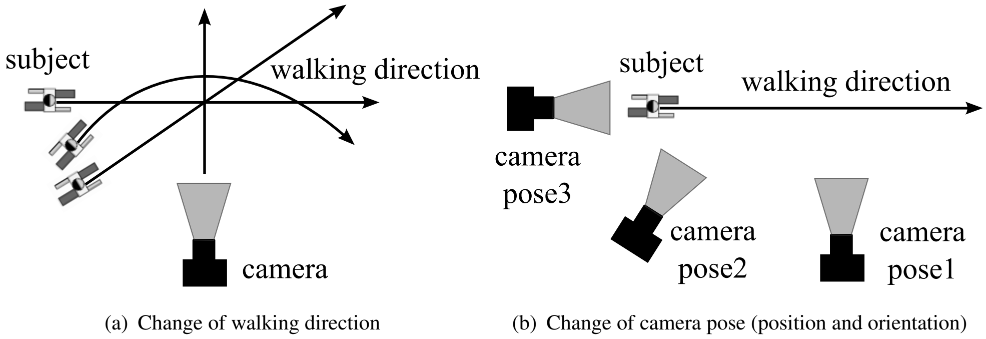
Figure 1.1: Silhouette images under different views from our collected dataset.

In realistic situations, people's gait can be captured from any camera viewpoint. The walking trajectory cannot be limited to a certain direction as people sometimes want to turn corners or avoid obstacles. There is also a varying distance between the subject and the camera. The problems associated with these observed walking conditions will be studied in this thesis. The following questions will be explored,

1. Can acceptable recognition performance be obtained when the subject's gait is captured from an arbitrary view?
2. Does the distance between the subject and the camera affect the appearance?
3. What influence does the shape of the walking trajectory (i.e. straight, curved) have on gait recognition performance?

1.1 Motivation and aims

View variation can happen in two situations. The first situation is when the position and orientation of a captured camera are changed. The second situation is when a subject changes his/her walking direction. These situations are illustrated in Fig. 1.2. To cope with view variation problems, several approaches have been proposed as will be explained in chapter 2. However, most of these approaches implicitly assume that people walk along straight lines and the view variation is relatively small in one gait cycle. Fig. 1.3 shows how appearance changes due to the walking direction changes at each frame along one gait cycle when the subject walks along a curved trajectory.



(a) Change of walking direction

(b) Change of camera pose (position and orientation)

Figure 1.2: Situations of view variation.

A theoretical study in [2] showed that there is a view variation in one gait cycle even if the subject walks along a straight line. This depends on a number of observation conditions, such as the distance between the subject and the camera and the viewing angle of a camera. The observed view variation is prominent when the distance between the camera and the subject is small and gait images are captured from an approximately side-view while it is small when the distance is large and gait images are captured from an approximately front-view. In a case of a large view variation in one gait cycle, extracted gait features are unstable as will be illustrated in chapter 3.

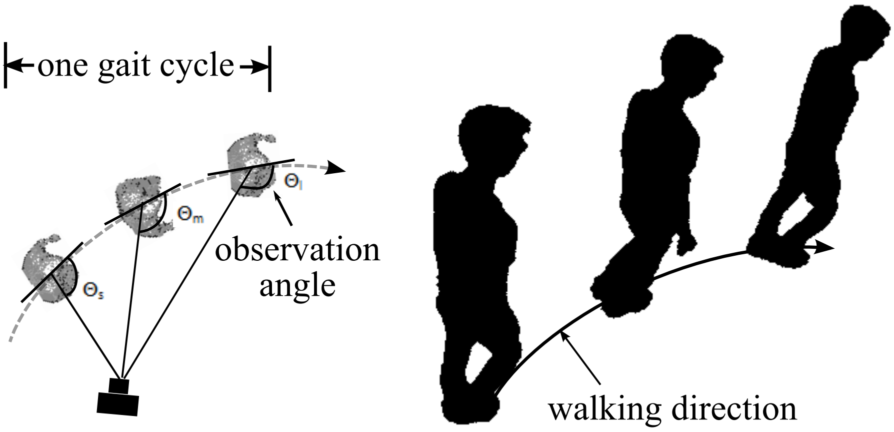


Figure 1.3: How the appearance and walking direction change in one gait cycle during walking along a curved trajectory.

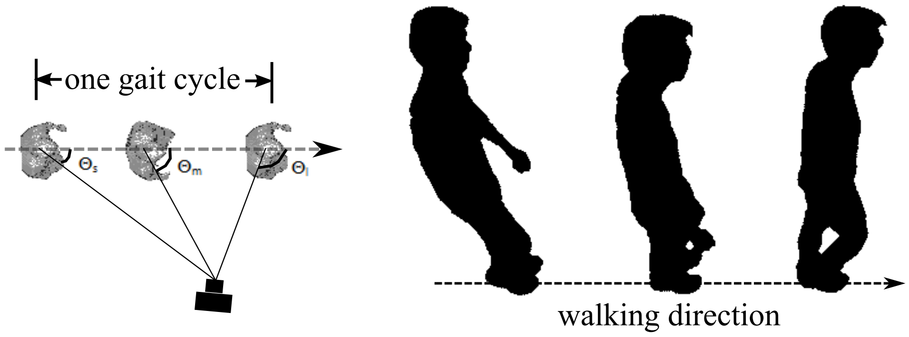


Figure 1.4: View variation in one gait cycle when the camera observes a subject from the side.

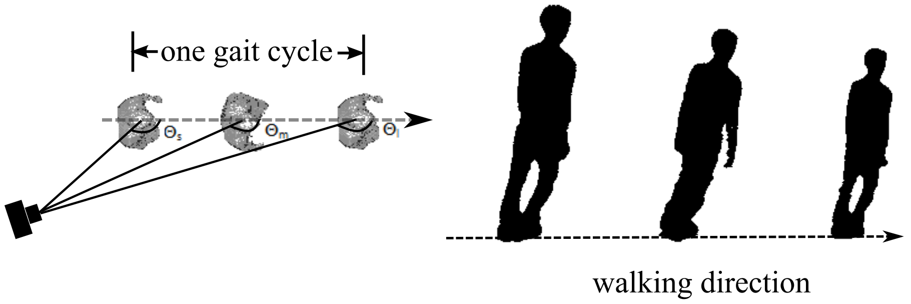


Figure 1.5: View variation in one gait cycle when the camera observes a subject from an approximately rear-view.

For walking along a straight line, Figs. 1.4 and 1.5 show several gait images from one gait cycle. The images are captured from an approximate side- and rear-view respectively. The difference between observation angles [1] in the start and end frame in one gait cycle can be clearly seen when the camera observes a subject from an approximately side-view. One of the effective solutions to the problem of view variation is to use 3D volumetric data since synthetic images from any viewpoint can be generated and used to match gait images from an arbitrary view. In order to

[1] observation angle is the angle between the direction of the walking of the subject and the direction of the camera with respect to the subject as illustrated in Fig. 1.3.

build 3D volumetric data of the subject, we need a synchronised multi-camera system to capture gait images from different viewpoints simultaneously. However, it would be impractical to use a multi-camera setup for both enrolment and recognition purposes. To exploit the full benefit of using the 3D volumetric data, a multi-camera system can be used at the enrolment stage only to build 3D reconstructions of people which can then be utilised to recognise gait images of subjects from a single camera(s).

1.2 Biometrics and gait

There is a wide range of biometrics used for identifying people based on their physiological or behavioural features. Physiological biometric refers to an individual's physical feature that remains relatively unchanged (e.g. fingerprints, iris patterns, DNA, facial expressions). On the other hand, behavioural biometric refers to a unique behaviour of an individual and thus it varies from one individual to another (e.g. signature, voice). Generally, the ideal biometric should have several important characteristics [46]: it must exist in each person (**universal**); it should not be shared by any two individuals (**unique**); it should not change with the time (**permanent**) and it should be easily captured and measured (**collectable**). Unfortunately, no ideal biometric exists, instead, a suitable biometric must be chosen according to a specific application.

Biometrics have already been used for human identification in many applications such as forensics, security and surveillance. Fingerprinting, which uses digital scans of an individual's finger to record his/her unique characteristics, is one of the widely used biometric technologies in forensics; although the intrusive nature and lack of acceptability by the public are the most problematic issues associated with this technology [47]. In surveillance and security scenarios, recognising people without their knowledge or cooperation is the key requirement and the use of recognition systems based on non-contact biometrics (e.g. facial expressions, ear patterns, gait) can be appropriate. In situations where the recognition has to be done within a large area and at a distance, the low resolution-image captured by a remote camera makes facial expressions and ear patterns useless while the use of an individual's gait for recognition can provide the most suitable choice. Gait uses a large proportion of peoples' bodies as well as their movements and thus it is hard to conceal or disguise. A normal video camera can be used to record an individual's gait in public areas without his/her intervention or knowledge. These attractive features give the priority for gait to be used in surveillance applications [100].

Gait can belong to both physiological and behavioural biometrics as the recognition can be based on the human shape as well as on motion. An early medical study [85] analysed a human's gait into 24 components by distinguishing the coordinated patterns of movements of different body parts. A standard pattern of motion was produced to recognise pathologically abnormal patients by attaching reflective targets to certain anatomical landmarks and using a strobe light flashing for illumination. This study suggested the uniqueness of the gait if all gait components are considered. A similar observation was found in biomechanics studies, which refered to the possibility of recognising people by their gaits when the patterns of their walking are measured in a repeatable and characteristic manner [120]. Inspired by these studies, biometrics researchers have been developing automatic gait recognition based on computer vision algorithms. Niyogi and Adelson [91] and Guo et al. [33] were the first to propose gait analysis techniques to identify people.

The potential of gait as biometric received more interest after the American Defence Advanced Research Projects Agency established and supported the Human ID at a Distance research program. The objective of this program was to motivate and encourage research on gait and other non-contact biometrics that can be used in surveillance technologies. This program also concentrated on extending the gait analysis study using small populations within constrained laboratory environments to real-world environments with large-scale populations. The program achieved many of its initial goals [66].

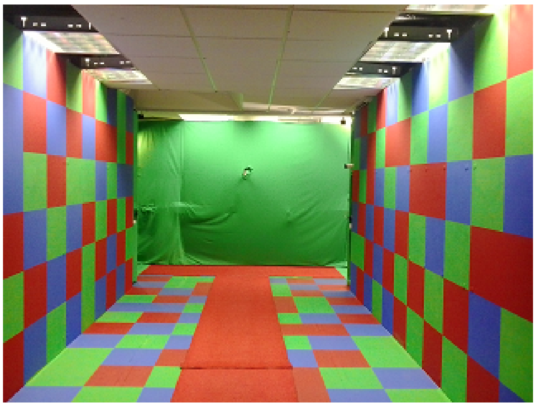


Figure 1.6: The Biometrics Tunnel from a frontal-view

1.3 Biometrics Tunnel

The Biometrics Tunnel at Southampton University represents a constrained environment, which is able to capture a variety of non-contact biometrics automatically as the subject walks through it. The design of the Tunnel is appropriate for high throughput environments such as airports. Fig. 1.6 shows the Tunnel from the entrance where there is a red carpet in the middle. The layout of the Tunnel was designed to mimic a narrow walking path with walls on both sides to constrain the walking direction of people. The Tunnel has a network of synchronised multi-cameras to record multi-view image sequences of people simultaneously. The synchronous multi-view sequences are used to build 3D volumes of people to solve the problem of view-dependency. The largest dataset collected by the Tunnel system includes over 200 subjects and contains gait, face and ear. We used the Tunnel as a tool to investigate the performance of matching 3D volumetric data against gait images captured by single cameras. The recognition performance using a subset from the largest gait dataset collected in the Tunnel showed that an average recognition rate of 97% was obtained when the matching was done against the far cameras near both ends of the Tunnel while an average recognition rate of 42% was recorded for the middle cameras (close to walking subjects). Further details about the matching process are given in chapter 3.

1.4 Thesis contributions

A wide range of gait-based human identification approaches has been developed to solve the problem of view variations. However, the majority of the work assumed that the appearance and orientation of a subject remain unchanged in a single gait cycle. They also assumed that people walk only along straight lines. In this thesis, we explored these two issues by enrolling people using 3D measurements while matching is performed against gait images captured by a single camera under any view. We describe the main contributions of this thesis in the following points:

- We conducted a set of experiments to illustrate the potential of matching a 3D volumetric data against 2D gait images from single cameras inside the Biometric Tunnel using the Gait Energy Image (GEI) as gait features. The results showed a low recognition performance when the matching was performed against the middle camera(s) due to a large view variation in a single gait cycle and omitting one of these cameras for the matching purpose and a high performance when gait images from the far-ends cameras were used for the matching. These results confirmed the theoretical study in [2] and showed that the distance of the subject from the captured camera influences the recognition performance. They also demonstrated that gait features are unstable when the distance between the camera and the subject is small and approximately side-view images are captured.
- We expanded and modified the layout of the Biometric Tunnel at Southampton University to produce a system capable of capturing synchronous multi-view and asynchronous (independent) single-view gait sequences. The main problems associated with the old Tunnel system, such as synchronisation, illumination and alignment, were addressed and resolved, which resulted in reliable and better silhouette extraction and 3D reconstruction results. Two asynchronous cameras were added to provide gait images from two independent views. Using the expanded Tunnel system, we collected two gait datasets including 3D and independent 2D gait sequences. The first gait dataset (small Soton 3D-2D gait dataset), which includes 17 people, was dedicated to evaluating the performance of recognising gait images of people walking along the central line in the Tunnel and captured by a perspective camera(s). The second dataset (large Soton 3D-2D gait dataset), which contains 50 people, was used to measure the gait performance of walking along straight or curved trajectories. Gait images from the independent cameras have a large view variation and an orientation change in a single gait cycle due to the wide-angle lens of these cameras and the closer distance to the walking subjects. The primary aim of adding these cameras was to provide a statistical view independence for the probe set.
- In chapter 5, we proposed a new gait recognition technique for identifying gait images with a high variation in appearance and orientation and captured by a camera mounted

on the wall in a narrow corridor. In this technique, each pair of 3D and 2D gait cycles in the gallery and probe sets respectively was matched by estimating the position of the 3D and back-projected 2D gait cycle in 3D space. The position of the 3D gait cycle was calculated from the centroid of the first volume while the position of the back-projected 2D gait cycle was estimated by back-projecting the first silhouette onto a walking plane of people in the gallery. After that, the centroid of the back-projected silhouette is calculated as a position of the back-projected 2D gait cycle in 3D space. Each volume in the 3D gait cycle was aligned according to the position of the back-projected 2D gait cycle in 3D space. The resulting synthetic silhouettes from the aligned 3D volumes share the same appearance, orientation and observation angle with the real silhouettes from the recorded camera. We calculate the Generic Fourier Descriptors (GFDs) from each synthetic and real silhouette along one gait cycle as gait features and then use Dynamic Time Warping for matching. We carried out extensive experiments to illustrate the effectiveness of the proposed technique on the small Soton 3D-2D gait dataset. The results clarified that a high recognition rate was achieved using a different number of features, truncated gait cycle and noisy probe silhouettes. The technique is also insensitive to perspective distortion from multiple viewpoints but it is sensitive to camera calibration errors.

- Although the previous technique copes with a wide range of view variation, it is not applicable when the walking direction is parallel to the optical axis of the camera (i.e. front/rear view) and its performance decreases when the people do not walk on a straight line because the 3D alignment is based only on the starting position of the gait cycle. To overcome these limitations, the second technique is proposed in chapter 6. In this technique, the 3D alignment is based on the foot positions in the three key frames at the start, middle and end of a gait cycle. The technique assumes the motion between these key frames to be linear. Based on this, the intermediate foot positions and local walking directions were calculated to cope with the problem of walking direction changes. Furthermore, in a 2D silhouette image the foot points of a subject is back-projected onto the floor plane to calculate their positions in 3D space. After that, the mean value of the back-projected points is calculated as the foot position of the subject in 3D space. This makes the technique completely view-independent. Based on the foot positions and walking directions, each 3D volume along one gait cycle was aligned using 3D rotation and translation respectively and then projected onto the 2D image plane to produce a synthetic image. Finally, subjects were identified using the GFDs as gait features. This technique was validated on two different datasets (the large Soton 3D-2D gait dataset and the Kyushu University 4D Gait Database) which include people walking along straight and curved trajectories. The results revealed that the performance of matching straight with straight walking was high and matching straight with curved walking was worse.

- Based on the results of the second proposed technique, we hypothesised that the discrepancy of gait patterns of people walking along different trajectories could influence the discriminatory power of features that are derived from the shape information only (i.e.

GFDs). Therefore, we explored the performance of gait features that are sensitive to the motion as well as the shape for gait recognition on straight and curved trajectories. We used the normalised width of the outer contour of silhouette and frieze patterns over four directions in each images as gait features. To calculate these features, we aligned each volume in a 3D gait cycle according to the position and walking direction of the corresponding silhouette in a 2D gait cycle. Then, the feature extracted from one direction in each probe frame was compared against those extracted from the nearest directions and from the neighbouring gallery frames within a sliding window. Using the sum rule to combine the matching results of the individual features from multiple directions, an improved recognition performance was achieved for matching straight with curved walking on the large Soton and the KY4D datasets.

The primary application of the work done in this thesis is in a forensic scenario where video footage of an unknown perpetrator captured by a CCTV camera is used for identity verification. To do that, gait patterns derived from the footage can be compared against those derived from the 3D volumetric data of the suspect walking in a controlled scenario (i.e. Biometric Tunnel) after projecting them onto a 2D image plane according to the pose of the perpetrator. The match can then be calculated according to the individual comparisons and used as an evidence based on the intra- and inter-subject variation. In fact, some work has already been done to use gait biometrics in forensic applications [11, 41, 78]. The following publications are related to the work done in this thesis:

- Fatimah Abdulsattar and John Carter. ***Performance Analysis of Gait Recognition with Large Perspective Distortion***. Accepted to be published in the IEEE International Conference on Identity, Security and Behaviour Analysis, Japan, 2016. The work in this publication is based on the first proposed technique in chapter 5.
- Fatimah Abdulsattar and John Carter. ***A Practical Technique for Gait Recognition on Curved and Straight Trajectories***. Accepted to be published in the 9th IAPR International Conference on Biometrics, Sweden, 2016. The work in this publication is based on the second proposed technique in chapter 6.

1.5 Thesis outline

The rest of this thesis is organised as follows. Chapter 2 describes the publicly available gait datasets and gives an extensive explanation of several gait analysis techniques with their main categories. In Chapter 3, we first illustrate the process of reconstructing 3D volumetric data from synchronous multi-view gait sequences captured inside the Biometric Tunnel. Then, we investigate the performance of matching 3D volumetric data against gait sequences from single cameras using one of the most poplar gait representations. Chapter 4 describes the main modifications and extension to the layout of the Biometrics Tunnel to collect synchronous multi-view and independent single-view gait sequences. In chapter 5, we present the first proposed gait recognition technique for highly distorted gait images captured by an arbitrary camera mounted on the wall in a narrow corridor. Chapter 6 explains the second proposed gait recognition technique for identifying people walking along straight and curved trajectories. In chapter 7, we explore the performance of gait features that are sensitive to the motion as well as the shape for gait recognition on straight and curved trajectories. Finally, a brief discussion about the conclusions and recommendations for future works are given in chapter 8.

Chapter 2

Human Recognition by the Way of Walking

Recognising people by their gait has received increasing attention by the research community over the last decades. One of the early results published in this field was by Johansson [49], who showed the possibility of discriminating human locomotion and other motion patterns using point light displays. Then, Cutting and Kozlowski [24] developed this capability to demonstrate that people could recognise their friends by their gaits. After that, several computer-vision based gait analysis techniques have been proposed since the beginning of 1990s. The research in this field started with small datasets of tens of people captured by single cameras and showed that a high identification rate could be achieved. Later, larger gait datasets of more than 100 people were captured using one or multiple cameras. These gait datasets also include a wide range of covariate conditions (e.g. clothes, carrying conditions, viewing angles, surface type, capturing environments, shoes, time elapsed) to enable reliable performance evaluation. The relative size of the available gait datasets is small (from about 10 to just over 300 subjects). This could be due to the lack of sufficient memory and processing time requirements for this type of biometric trait. With the progress of computer technologies, a larger gait dataset of more than 4000 subjects has been collected and a high identification rate of 97.5% [42] has been achieved. Recently, Makihara et al. [79] from Osaka University collected 47,615 subjects over 246 days indoors to produce the largest gait dataset in the world. A few scientific institutions have also 3D gait datasets and some gait analysis techniques have been developed based on 3D data. The main advantage of 3D data is that it solves the problem of view dependency.

Generally, three scenarios for gait recognition can be specified. Several gait analysis techniques have been proposed to deal with each scenario. The first scenario is called **single-view gait recognition** where both gallery and probe samples are from the same view. The performance of single-view gait recognition techniques is significantly influenced by view variation as the derived gait features from a specific view would be altered under another view. The second scenario is when the gallery and probe samples are from different views. This scenario is called **cross-view**

gait recognition. The third scenario is named **multi-view gait recognition** where the gallery samples are from multiple views and the probe samples are from single or multiple views. The techniques proposed for the second and third scenarios can better deal with the problem of view variation than those for the first scenario.

The first part of this chapter talks about gait dataset types (i.e. 2D and 3D gait datasets). For each type, several datasets have been described in detail, in terms of the number of subjects, capturing environment, number of recording cameras and covariate conditions. The second part discusses several gait analysis techniques proposed for the gait recognition scenarios. Features that describe the pattern of walking and discriminate different people based on their gait are mainly divided into model-based and holistic-based. These two types of gait features are also covered in this chapter.

2.1 Gait datasets

Several scientific institutions have been recording and developing their own gait datasets in order to fairly compare and evaluate the performance of gait analysis techniques. Generally, we can divide the available gait datasets into two main classes: 2D and 3D datasets. The 2D gait datasets include single-view gait sequences that are captured by a single camera and multi-view gait sequences that are captured using multiple cameras without having time synchronisation or calibration information. On the other hand, the 3D gait datasets contain 3D volumetric data that are reconstructed from synchronous multi-view sequences using calibration information about the cameras. The following subsections cover these two types of gait datasets.

2.1.1 2D Gait datasets

Most of the early gait datasets, such as Southampton (Small Soton) [20], South Florida (UCSD) [67], Georgia Institute of Technology (GaTech) [9] and the University of Maryland dataset [22], were collected using a single camera and filmed either indoors or outdoors. Other datasets were recorded using multiple cameras. One of the popular available 2D multi-view gait datasets is the Motion of Body (CMU MoBo) [31], which was collected by Gross and Shi from Carnegie Mellon University in 2001. This dataset includes 25 subjects walking on a treadmill indoors. Six high resolution colour cameras distributed around the treadmill were used to film the subjects in four different scenarios: slow walk, fast walk, incline walk and walking carrying a ball. Although, the cameras used to capture this dataset were synchronised, no calibration information is provided to reconstruct 3D gait sequences. An example of the silhouettes from the six views in the CMU MoBo dataset is shown in Fig. 2.1.

The Gait Challenge dataset (USF) [95, 71], which is one of the most widely used gait datasets, was released by the University of Florida with the cooperation of Notre Dame University. This dataset includes video sequences of people walking outdoors in two elliptical paths under varying

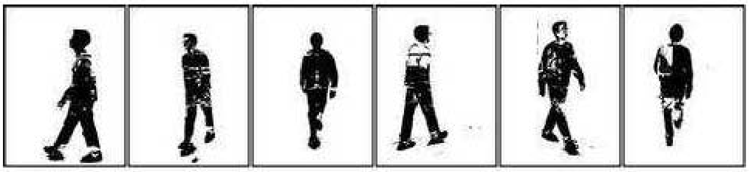


Figure 2.1: Silhouettes from different views in the CMU MoBo dataset [31].

covariate factors. These factors include different viewing angles, shoe types, walking surfaces, baggage and time elapse. Two cameras were used to record subjects walking one time through the field of view for different combinations of covariate factors. This dataset initially includes 74 subjects, but this number was extended later to 122 subjects [97]. The males constitute 75 percent of this database, whilst the female only 25 percent. Fig. 2.2 shows a subject walking on a concrete and on a grass surface from two different viewpoints.

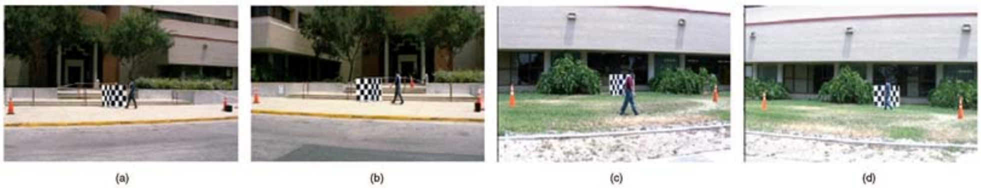


Figure 2.2: A subject from the Gait Challenge dataset captured by (a) the left camera for a concrete surface, (b) the right camera for a concrete surface, (c) the left camera for a grass surface and (d) the right camera for a grass surface [97].

The Pattern Recognition National Laboratory in the Institute of Automation, Chinese Academy of Sciences collected different versions of the CASIA gait dataset. The first version (CASIA-A) [119] consists of twenty subjects recorded outdoors. Each subject walked four times along a straight line in each of three different viewing angles relative to the camera image plane, lateral (90°), oblique (45°) and front (0°). A total of 240 gait sequences were filmed over two different days using a single digital camera mounted on a tripod. The second version (CASIA-B) [123] was then released, which contains 124 subjects filmed indoors using eleven USB cameras distributed evenly around the left-hand side of the walking subjects as shown in Fig. 2.3. Each subject walked ten times under different clothing and carrying conditions. The number of males in this dataset is 93 whereas the number of females is only 31. This dataset is frequently used by the researchers to evaluate the performance of multi-view gait analysis techniques. In 2005, another version of the CASIA dataset was also produced, namely CASIA-C [110]. This dataset includes 153 subjects (130 males and 23 females) filmed outdoors at night. A thermal infrared camera was used to capture the subjects from a side-view. Each subject walked ten times under four different conditions: slow walk, normal walk, fast walk and carrying a bag.

The university of Southampton introduced one of the most comprehensive datasets called the Human ID at a Distance (HID) dataset [102, 89]. This dataset includes over one hundred subjects filmed indoors and outdoors. For the indoor data, the subjects walked along a straight line and on a treadmill. The same subjects were also filmed outdoors in a busy environment to simulate a real

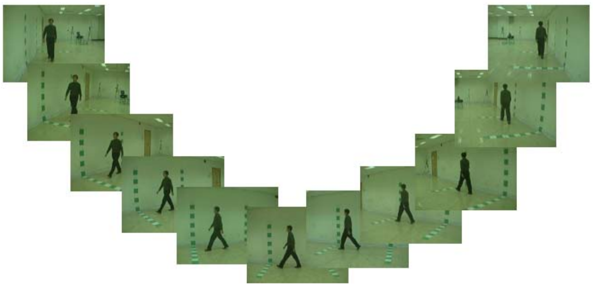
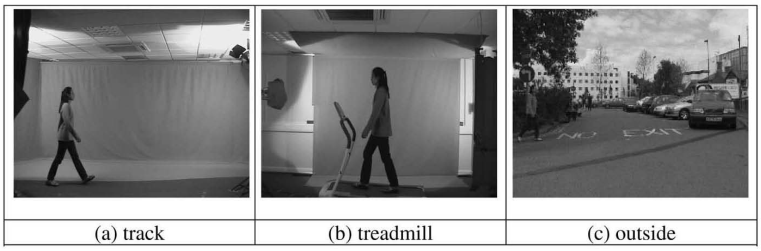


Figure 2.3: Samples from the CASIA-B dataset from 11 viewpoints [123].

scenario. Two cameras were used to film the subjects from two different viewpoints: side and oblique. Fig. 2.4 shows frames from the Southampton University HID dataset captured indoors and outdoors.



(a) track (b) treadmill (c) outside

Figure 2.4: Samples from the Southampton HID dataset in three different situations [88].

Due to the promising results in the field of gait recognition during the previous years, new gait datasets have recently emerged. The researchers at Osaka University collected the OU-ISIR large population gait dataset (OULP-CIVI) [42], which comprises 4007 subjects. This dataset includes a sufficient number of subjects for each gender (2135 males and 1872 females) and also covers a variety of age groups, ranging between 1 and 94 years. Four sequences were captured for each subject using two cameras positioned at approximately $4m$ from the walking path such that the first camera captured a transition of views from a frontal oblique to a side image while the second camera captured a range of views from a side to a rear oblique. The sufficient number of subjects involved in each age group as well as the unbiased representation of males and females enable the researchers to analyse the variation of gait recognition performance according to gender and age group. Osaka University developed another dataset [80] to provide a reliable performance evaluation based on within-subject variation. This dataset includes 34 subjects walking at varying speeds ranging from 2km/h to 10km/h, 68 subjects with a combination of 32 types of clothes, and 200 subjects captured by 25 synchronous cameras. It focuses on subjects walking on a treadmill and also includes a wide variation of age and gender. Due to the recent development in computer technology, Osaka University collected a dataset of 47,615 subjects over 246 days by running an exhibition in association with a science museum [79]. In this dataset, a seven networked cameras of 1280×960 pixels at 25 frames per second had been placed at intervals of $15°$ along a quarter of a circle whose centre was set up at the centre of the walking path except the front/rear camera (i.e. Cam 7) where its position was on the opposite side of the other cameras as shown in Fig. 2.5. The radius of the circle was about $8m$ and the cameras were at a height of approximately $5m$. Each participant walked one time with carrying condition and 2 times normally for each of the outward and return directions.

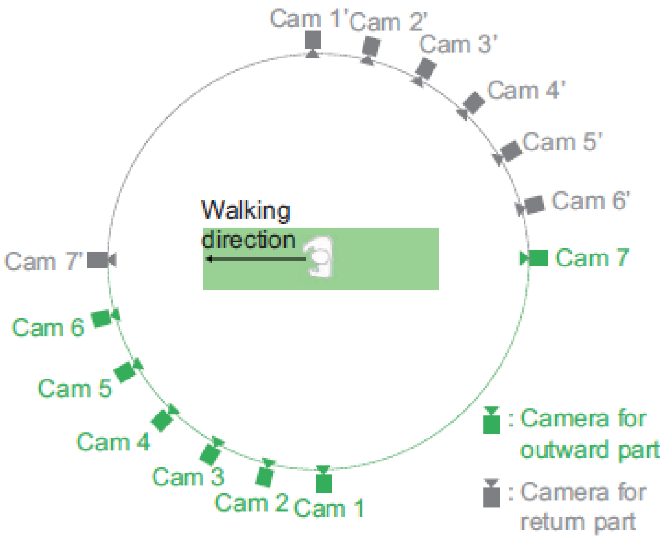


Figure 2.5: Camera setup in the largest scale dataset [79].

Unlike other gait datasets, Hofmann et al. [39] introduced the TUM Gait from Audio, Image and Depth (GAID) dataset, which comprises RGB image, depth and audio. A Microsoft Kinect sensor was used in this dataset to record 305 subjects (186 male and 119 female) over two sessions in Munich, Germany. This sensor was able to simultaneously provide a typical video stream, a depth stream as well as a four-channel audio. Each subject walked six times normally, two times carrying a backpack, and two times with varying shoes. Fig. 2.6 demonstrates subjects walking in a normal way and with covariate variations.

2.1.2 3D Gait datasets

Few researchers have collected 3D gait datasets to evaluate the performance of gait recognition algorithms based on 3D data. Seely [98] collected a large multimodal (gait, face and ear) dataset, which includes over 200 subjects and 2000 samples. This dataset was recorded indoors using 14 cameras: 12 cameras to capture synchronous multi-view gait sequences, and the remaining 2 cameras to record a face video and an ear imagery. Later, Matovski et al. [82] captured a new multimodal (gait and face) temporal dataset to analyse the effect of time on gait recognition performance. This dataset contains 25 subjects (17 male and 8 female) recorded by 12 synchronous cameras. The subjects were filmed in four sessions over 9 months.

The previously described gait datasets assume that people walk only along a straight line. However, walking paths in our daily lives tend to be curves rather than straight lines [108]. Two gait datasets have been recently captured from multiple views, including people walking on

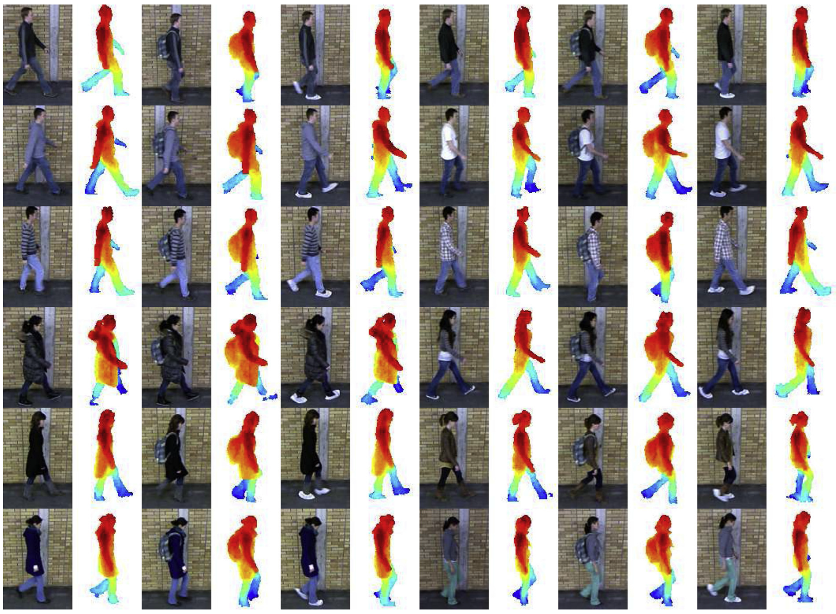


Figure 2.6: Several samples from the TUM dataset in different covariate conditions [39].

straight lines as well as curved trajectories. The calibration information of the captured cameras in these datasets was also provided to allow the building of 3D data. López-Fernández et al. [75] collected the AVA Multi-View Dataset for Gait Recognition (AVAMVG). This dataset consists of twenty people (4 females and 16 males) walking along ten different trajectories in an indoor environment. Three of these trajectories are straight lines whereas the remaining trajectories are curves. Video sequences of the walking subjects were captured by six synchronous colour cameras distributed around the scene at a height of 2.3m from the floor and directed downward. The video was captured with a resolution of 640 × 480 pixels and at a rate of 25 frames/second. The body parts of the people are partially visible from two particular camera viewpoints most of the time more than the remaining four cameras [15]. Fig. 2.7 shows an example for samples with partially visible body parts from two different viewpoints. The illumination was provided using natural light that entered the recording scene through four windows to provide a real scenario. Iwashita et al. [44] published the Kyushu University 4D gait Database (KY4D). This dataset includes 42 people recorded by 16 synchronous cameras distributed around a circular studio of radius 3.5m. These cameras are arranged at two different heights. Fig. 2.8 shows an example for samples from different viewpoints. The video was recorded at 20 frames/second and with a resolution of 1032 × 776 pixels. Each subject walked four times in a straight line and once along two curved trajectories of radius 3m and 1.5m. This work was carried out independently and in parallel to the work described in this thesis.

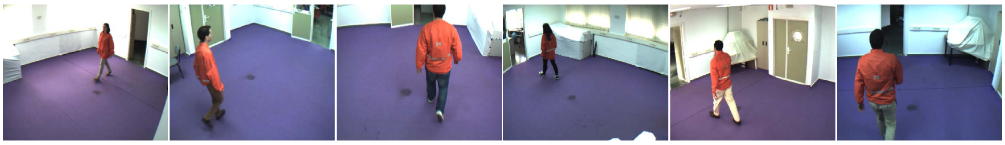


Figure 2.7: Partially visible body parts from the 3rd and 6th viewpoints in the AVAMVG dataset [15].

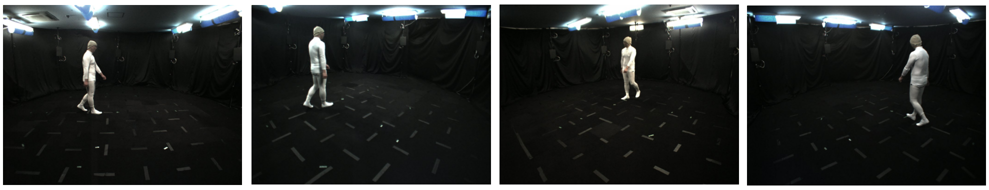


Figure 2.8: Samples from the KY4D gait dataset [44].

2.2 Single-view gait analysis techniques

The majority of the existing techniques for gait recognition are regarded as two-dimensional (2D). These techniques deal with video sequences captured by only a single camera (single-view gait recognition) and their main advantages are the requirements of less computational costs and small storage capacities. However, they suffer from a view-dependency problem. According to the method of feature extraction, these techniques can be divided into two groups: model-based and holistic-based. Each of these groups has its benefits and drawbacks. The following two subsections describe these two groups in more detail.

2.2.1 Model-based gait analysis techniques

The techniques in this group depend on fitting a model to a series of images by tracking different body parts such as arms and legs. The parameters of this model are used to derive gait feature (signatures) for identification. Basically, there are two models used in these techniques: structural and motion. A structural model involves a set of static features that describe different body parts such as human body height or stride length. This model uses a prior information for the human body structure to build up primitive shapes such as stick figures, 2D contour or cylinder models as shown in Fig. 2.9. On the other hand, a motion model describes a set of dynamic features (kinematic information) such as joint angles trajectories. The main advantages of model-based approaches are their abilities to handle occlusion, noise and a small view variation. These advantages are crucial for practical applications. However, these techniques require high-quality image sequences and extensive computation to derive high accuracy model parameters [66, 90].

BenAbdelkader et al. [6] extracted stride and cadence as gait features by assuming that people were walking on a known plane with a constant velocity and that the captured camera was calibrated. Using a dataset of 17 people with 8 samples each, an identification rate of 40% was obtained. Lee and Grimson [64] fitted ellipses to seven local regions in a silhouette, representing different human body parts. The ellipse parameters from these regions were used to identify an individual. This technique achieved a high recognition rate for side-view gait sequences. However, its major limitation was the use of fixed region boundaries. Cunado et al. [21] fitted a stick model to a sequence of images using Genetic Algorithm based Velocity Hough Transform. A gait signature was extracted using the Fourier components of the signal related to the motion of the upper leg. The results demonstrated that this method can handle a high level of occlusion, which is essential in gait. Yam et al. [122] extended the work of Cunado by constructing a structure and motion model of legs to recognise walking and running subjects. This algorithm was tested on a dataset of 20 people walking and running on a treadmill. The recognition results showed that running has a greater discriminatory capability than walking with a recognition rate exceeding 90%. Wagg and Nixon [116] employed an advanced model in which a pair of ellipses are used to represent the head and torso and two pairs of line segments are used to represent each leg. Progressive stages of fitting are used to find the parameters of the model as gait features. This approach was evaluated on a large dataset of 115 subjects and obtained a recognition rate of 84%. Recently, Bouchrika et al. [12] proposed a new model to describe the motion of the joints using Elliptic Fourier Descriptors. The model was fitted to a sequence of images using the Hough Transform. This algorithm recorded a recognition rate of 92% on a dataset of 120 people.

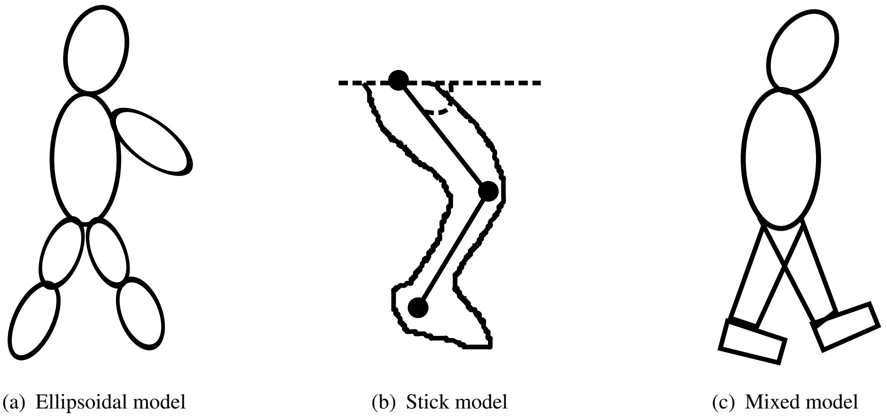

(a) Ellipsoidal model
(b) Stick model
(c) Mixed model

Figure 2.9: Primitive shapes used in model-based approaches, (a) ellipses fitting in silhouette parts [64], (b) stick model fitted in leg region [21] and (c) model consisting of a combination of shapes [116].

2.2.2 Holistic-based gait analysis techniques

Unlike model-based techniques, holistic-based techniques derive a gait signature directly from image sequences without modelling human body parts. These techniques are insensitive to the quality of the image sequence and require less computational costs in comparison with model-based techniques [66]. However, they are sensitive to view variations. Gait biometrics researchers proposed several holistic-based approaches, depending on either analysing the pattern of motion or describing the moving shape over a sequence of frames.

BenAbdelkader et al. [7] proposed one of the simplest holistic-based approaches by computing the correlation between each pair of silhouettes in a sequence to produce a self-similarity plot for individual recognition. Using a dataset of 40 sequences obtained from 6 subjects, this method obtained an identification rate of 93%. Philips et al. [95] introduced a baseline algorithm, in which the probe sequence was divided into a number of subsequences. Then, each of these subsequences was correlated with the gallery sequence. Finally, the similarity was computed as the median value of the maximum correlation between the gallery sequence and each of these probe subsequences. The identification rates obtained by this algorithm on the Gait Challenge dataset (USF) were 79%, 66% and 29% for a variation of views, shoes and surfaces respectively. Despite its simplicity, this algorithm was impractical due to excessive computational requirements. Collins et al. [17] proposed a more practical technique using a key frame analysis. They estimated one gait cycle from each test sequence to determine key frames. These frames were then compared against training frames using a normalised correlation. This technique was evaluated on four datasets with different covariate conditions. The results showed that the technique has a good recognition performance under a fixed viewing angle and can handle noisy silhouettes. Sundaresan et al. [109] introduced a statistical method based on Hidden Markov Model (HMM) to recognise individuals walking in a fronto-parallel pose using a compact representation. This method produced better results than the baseline algorithm [95].

Several techniques used a single gray-scale image to provide a compact representation for gait over a complete gait cycle. Liu and Sarkar [72] presented a simple yet robust representation for gait recognition, called average silhouette. In this approach, the silhouettes were normalised according to their heights, aligned horizontally and then averaged over one gait cycle. This approach has a similar performance as the baseline algorithm on the Gait Challenge dataset. This representation was also used by Han and Bhanu [34] and Veres et al. [115]. Han and Bhanu [34] computed the Gait Energy Image (GEI) by averaging silhouettes in one gait cycle. Statistical analysis was then applied to learn effective features for classification. Veres et al. [115] measured the contribution of static and dynamic information for gait recognition performance using the average silhouette and a differential silhouette. Lam and Lee [61] suggested another 2D representation called Motion Silhouette Image (MSI) by computing a function of a historical motion for each pixel over all silhouettes in one gait cycle. Liu and Zheng [68] introduced an improved temporal template called Gait History Image (GHI), in which the static and dynamic characteristics of gait modelled in a comprehensive way. To mitigate the effects of covariate conditions that affect a body appearance, another representation called Gait Entropy Image (GEnI) [4] was developed by calculating the Shannon entropy for each pixel in the silhouettes over a complete gait cycle to capture most of the motion information. The Gait Flow Image (GFI) was also proposed by Lam et al. [60] where the magnitude of the optical flow was computed and averaged for the silhouettes in one gait cycle. Fig. 2.10 shows examples of different gait representations for human recognition. Hofmann and Rigoll [40] developed a new gait representation where Histograms of Orientated Gradient (HOG) was applied on each RGB image along one gait cycle. The final representation was computed by averaging the resulting gradient histograms. To improve the performance of this representation, all background pixels in each RGB image were set to zero before computing the histogram. The recognition results on the USF HumanID gait dataset demonstrated the effectiveness of this gait representation for human identification.

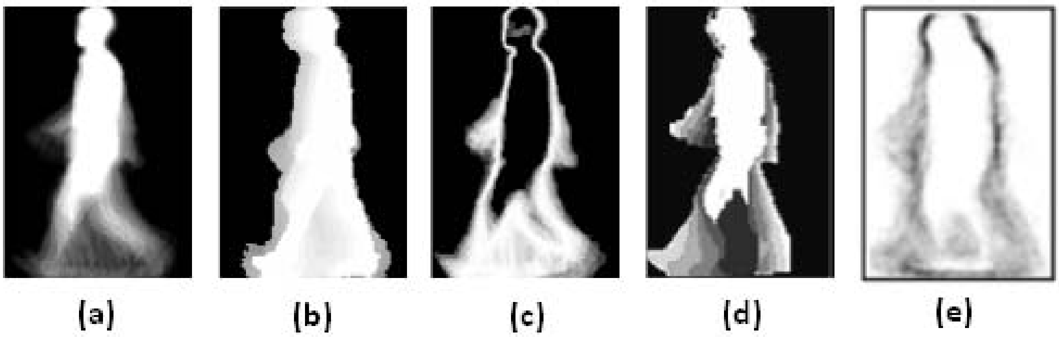


Figure 2.10: Various gait representations for holistic-based techniques: (a) GEI [4], (b) MSI [4], (c) GEnI [4], (d) GHI [68] and (e) GFI [60].

The variation of a periodic pattern in a silhouette sequence spatially and temporally provides a description for subjects and their movements, which could be exploited for recognition. Liu et al. [70] proposed frieze patterns by projecting a silhouette sequence horizontally and vertically to produce a pair of 2D images that are periodic with respect to the temporal domain. Lee et al. [65] introduced an extended version of frieze patterns using key frames subtraction. Kale et al. [52] extracted the width of the outer contour of silhouettes in one gait cycle as gait features. These features were then compared using Dynamic Time Warping. The efficacy of this method was demonstrated using three different gait datasets. Statistical techniques, such as moments, were also used to exploit the periodic variation in moving shapes over a sequence of frames by providing a set of features invariant to rotation, translation and scaling. Shutler et al. [104] developed new types of moments, called velocity moments, by incorporating velocity information of moving objects into the centralised moments [28]. These new moments provide discriminative properties useful for recognition. However, the major drawback of these moments was that the resulting features were highly correlated, due to the non-orthogonality of the original moments from which the features derived. To mitigate this weakness, Velocity Zernike Moments were later developed by Shulter and Nixon [103] to provide less correlated and more compact descriptions for human motion. These new moments depend on the orthogonal Zernike Moments. Foster et al. [29] used the temporal variation of an area inside a masked region over a sequence of silhouettes as a signature for gait recognition. A recognition rate of over 75% was obtained on the HID dataset, which consists of 114 subjects, when combining information from different masked regions.

Some techniques analysed the shape of a moving person directly to derive a pattern of motion for recognition. Hayfron-Acquah et al. [37] extracted the symmetry of motion from a silhouette sequence as a signature for gait recognition. The symmetry operator was used to measure the symmetry between silhouette points. The resulting symmetry images from a sequence of silhouettes were then averaged and Fourier coefficients were finally calculated as gait features for recognition. This technique was evaluated on a small dataset and a recognition rate of around 95% was recorded. Wang et al. [117] used a Procrustes shape analysis to encode the temporal variation of a silhouette's shape over one gait cycle. Each silhouette sequence was represented as a set of complex configurations in a common coordinate. The Procrustes mean shape was then calculated as a gait signature and compared using the Procrustes distance as a metric. A recognition rate of around 75% was obtained on the CASIA-A dataset. Tassone et al. [114] used the movements of boundary points in a silhouette sequence to build a temporal Point Distribution Model (temporal PDM). The resulting model was then used to discriminate subjects walking on a treadmill at varying speeds. A study in [27] captured geometric properties of the silhouettes boundaries by computing contour curvatures locally. This new gait feature was evaluated on the OU-ISIR Large Population dataset, which includes over 4000 people. The Rank-1 identification rate ranged from 91.6% to 93.3% for a list of viewing angles $(55^\circ, 65^\circ, 75^\circ, 85^\circ)$.

2.3 Multiple-view gait analysis techniques

The techniques described in the previous section can deal only with a single-view gait recognition. In real world scenarios, a subject can be recorded from several camera viewpoints and his/her appearance can vary in different views. This will affect the performance of gait analysis techniques that are designed for a specific view. To tackle the problem of view variation, several gait analysis techniques have been developed by assuming that the gait data in the gallery and probe can be from different views. These techniques can be divided into view-invariant and view transformation.

2.3.1 View-invariant gait analysis techniques

These techniques aim to extract gait features which are invariant to view variation. The view-invariant techniques are often used for a cross-view matching where the viewing angles of the gallery and probe are different. Kale et al. [51] introduced a simple technique to synthesise a side-view of the gait from any other arbitrary view by using a perspective projection in a sagittal plane where the subject was assumed to be a planar object and far enough from the camera. This technique produced a recognition rate of over 80% on a dataset of 12 subjects walking along arbitrary straight directions when the angle between the sagittal plane of the subject and the image plane is small. However, the performance deteriorated when the angle increased. Han et al. [35] used a statistical method to extract view-invariant features from the GEI by seeking for the overlap parts of the gait sequences from different views to build a representation for gait matching across views. A small dataset of 8 people walking across 10 different directions in an outdoor environment was collected to investigate the performance of this method and a good recognition rate was achieved when the view difference between the training and testing data was small but its performance degraded when the view difference was large (i.e. a little overlap between gait sequences). Jean et al. [48] introduced a method to normalise body part trajectories for people walking along different directions. A homography transformation of an observed walking plane was firstly computed from 2D trajectories of head and feet. This was then used to normalise body part trajectories from an arbitrary view to a side view. However, self-occlusion makes the process of tracking body parts unreliable. Goffredo et al. [30] estimated lower limb poses using a markerless motion estimation procedure. These poses were then reconstructed in a sagittal plane using a viewpoint rectification by assuming that the articulated leg motion is approximately planar. Next, the rectified poses were used to derive a set of view-invariant gait features. The performance of this technique was evaluated on a dataset of 65 subjects walking freely along different directions and a mean classification rate of 73.6% was obtained across all views. This technique works efficiently when a view variation is large but its performance drops when a subject is captured from the front.

2.3.2 View transformation gait analysis techniques

The second category relies on learning a mapping relationship of gait features observed under different views. Some techniques in this category construct a View Transformation Model (VTM) by building a matrix where each row represents gait features of different subjects under the same viewing angle, and each column represents gait features under different viewing angles for the same subject. Singular Value Decomposition (SVD) is then used to factorise this matrix into two independent sub-matrices: view independent and subject independent. These sub-matrices represent the basis for VTM construction. Suppose there are N subjects and Z viewing angles, the factorization process is done as follows [56]:

$$\begin{bmatrix} G_1^1 & \dots & G_1^N \\ \vdots & \ddots & \vdots \\ G_Z^1 & \dots & G_Z^N \end{bmatrix} = USV^T = \begin{bmatrix} P_1 \\ \vdots \\ P_Z \end{bmatrix} \begin{bmatrix} v^1 \dots v^N \end{bmatrix} \tag{2.1}$$

Where G_z^n is M dimensional feature vector for the n^{th} subject under the z^{th} viewing angle, U is the $ZM \times N$ orthogonal matrix, V is the $N \times N$ orthogonal matrix, S is the $N \times N$ diagonal matrix of singular values, P_z is the $M \times N$ submatrix of US, and v^n is the N dimensional column vector. Makihara et al. [81] computed a number of frequency gait features using Fourier analysis based on a gait period to build up a VTM. This was then used to project gallery gait features into the same viewing angle as that in a probe set before calculating the gait similarity for matching. Using an in-house dataset of 20 subjects and 24 walking directions, a recognition rate ranged between 10% and 90% depending on the view difference of gaits in the gallery and probe. To tackle the problems of over-fitting and over-sizing associated with VTMs, Kusakunniran et al. [56] used a truncated SVD (TSVD) for factorisation. In this technique, the GEI was extracted as a gait feature from each sequence and optimised using Linear Discriminant Analysis (LDA) to build a VTM. Using multi-view to one-view transformation, the technique achieved a better performance than [81] on the CASIA-B dataset. To efficiently deal with a partial occlusion in silhouettes and further improve the performance of view transformation, a VTM construction was considered as a regression problem in [57, 58, 59]. The regression concept was exploited to seek for correlated motions of gaits under different views. Multiple regression processes were used to train a VTM to estimate gait features from one viewing angle using correlated information in gait features from other viewing angle(s).

Instead of reconstructing gait features from different views, Bashir et al. [5] used Canonical Correlation Analysis (CCA) to model the correlation of gait features observed across different views by projecting them into learned subspaces which maximised their correlations. Then, the correlation strength based on the CCA was used to measure the similarity between gait features in these subspaces. To alleviate the computational cost of using the CCA with two sets of high-dimensional vectors, Xing et al. [121] proposed a novel approach called complete Canonical Correlation Analysis. In this approach, the traditional CCA was reformulated into two stable eigenvalue decompositions to avoid computing the inverse of a high-dimensional matrix for gait image data. The analysis on the CASIA-B dataset yielded a recognition rate of 20% when the view difference is 90° and 98% when the difference is 18°. Liu and Tan [69] learned multiple subspaces using the LDA to extract discriminant features from each view in the training data. A new gait descriptor called Radon Transform based Energy Image (REI) was calculated as gait features. Then, testing features were projected onto each of these subspaces to match against the training features in the same subspace. Finally, the matching results from all subspaces were fused to produce the final distance metric for recognition. Using three different views $(18^\circ, 108^\circ, 162^\circ)$ for training from the CASIA-B dataset, an average recognition rate of 90.69% and a minimum recognition rate of 84.68% were achieved over 11 testing views.

The techniques in this category are more efficient in dealing with large view changes when sufficient training data are available to the learning process, compared to view invariant techniques. Some of them can also be used for both cross-view and multi-view gait recognition. However, these techniques depend on a learning process and hence their performance will drop if they are required to recognise gait from an arbitrary (untrained) view. Recently, Muramatsu et al. [84] overcame this limitation by proposing an arbitrary VTM (AVTM) to recognise gaits from an arbitrary view. Under any view change, a corresponding VTM was constructed using training gait features generated from 3D human models of non-target subjects under any required views. The 3D gait information was only used to construct a VTM. This technique achieved a better accuracy than conventional VTM-based techniques on the treadmill data and large scale population data from the OU-ISIR gait dataset [80, 42] for cross-view matching. Using 100 subjects from the treadmill dataset, the AVTM recorded a range of recognition rates from about 20% to 100% depending on the view difference between the gallery and probe sets. On the other hand, the technique achieved recognition rates between almost 40% and 78% for the view differences of up to 30° using 1912 subjects from the large scale dataset. It should be noted that the generated image by the AVTM may not correspond to the real image of the subject as the 3D models of the target people were not used in the training data to learn the VTM.

2.4 3D Gait analysis techniques

The 3D techniques integrate information from multiple synchronous cameras to build a 3D gait model. As the movements of a human occur in 3D space, 3D gait data contain more information than 2D data. The 3D reconstructed data provide an efficient solution for the problem of view dependency since 2D gait sequences from any viewpoint can be synthesised. Lee [63] reconstructed 3D volumetric sequences to mitigate the problems of view dependency and self-occlusion of a 2D ellipse fitting approach proposed by Lee and Grimson in [64]. Four synchronous cameras were used to reconstruct 3D models of the walking subjects. A virtual camera was placed perpendicular to the estimated walking path at each time instance. Synthetic silhouettes were then generated from a fronto-parallel view. Using a dataset of 27 individuals and 225 sequences, a recognition rate of 67% was obtained using the synthetic silhouette sequences. Bodor et al. [10] also synthesised fronto-parallel view sequences from 3D models of people walking in any direction using several cameras mounted above and around a motion path. After that, gait features based on the Principal Component Analysis (PCA) were calculated from the resulting synthetic sequences and used for action classification. Sivapalan et al. [105] segmented a 3D model into four parts and fitted ellipsoids directly to these parts without extracting 2D silhouettes. A Fourier representation was then used to model gait features extracted from ellipsoid parameters for gait recognition. The recognition performance was evaluated using the multi-view CMU MoBo dataset, and an average recognition rate of 85% was recorded under different walking speeds and carrying item conditions.

Zhao et al. [125] proposed a method for fitting a 3D human model on gait sequences extracted from multiple cameras. A skeletal model was fitted to the first frame in all views by manually selecting the initial model parameters such as position, orientation and joint angles. The motion was then tracked on the subsequent frames using a local optimisation algorithm to estimate the trajectories of lower limbs. The static features represented the lengths of different body parts while the dynamic features represented the motion trajectories of lower limbs. Dynamic Time Warping was used for matching. A subset of 10 persons was randomly selected from the multi-view CMU MoBo dataset to evaluate the performance of this technique and a recognition rate of 70% was recorded when combining the static and dynamic features. Bhan and Han [8] proposed a highly complicated 3D kinematic model, consisting of more than thirty degrees of freedom, using gait sequences captured by only a single camera. Fitting such a model to the 2D silhouette sequences was a non-trivial task. Several assumptions were proposed to reduce the complexity of this model. Static and dynamic features were extracted from the human motion characteristics and body parts measurements respectively. Using a dataset of eight people walking in an outdoor environment, an optimum recognition rate of 83% was achieved using both static and dynamic features.

Several studies have been conducted using the 3D volumetric dataset captured by the Southampton Biometric Tunnel, multimodel 3D Soton dataset [98]. Initially, Seely et al. [100, 99] captured multi-view gait sequences for 103 subjects using 12 synchronous cameras and reconstructed their 3D volumes. Then, gait features were extracted by firstly projecting each volume into the three orthogonal viewpoints (side-on, front-on and top-down) to produce three sets of synthetic images and then computing the average silhouette for each synthetic sequence. A CCR of 99.6% was obtained by combining the features from the three orthogonal viewpoints. Ariyanto and Nixon [3] adopted a simple yet 3D model fitting approach in which a structural model consisting of articulated cylinders with 3D degrees of freedom at each joint was fitted into a 3D data to model the human lower leg. In this technique, a correlation filter and Dynamic Time Warping were used to fit the model into the data to extract both structural and dynamic gait features. Using 46 subjects from the multimodel 3D Soton dataset, this technique was able to achieve up to 79.4% recognition rate. Matovski et al. [82] analysed the effect of time on gait recognition performance. Several experiments were conducted for this purpose using the 3D Soton temporal dataset. In these experiments, the effect of time was isolated from other covariate factors that have a significant impact on performance such as clothing, footwear etc. Performance analyses showed a slight effect of the short-medium term of time on gait recognition performance. They also showed that clothing had a major impact on performance as the recognition rate dropped drastically when there was a significant change in clothing.

Alternatively, some techniques used point cloud data captured by depth sensor devices to build a partial 3D volumetric representation, called 2.5D human model. Tang et al. [113] introduced a new feature extraction method based on a 2.5D human model. Point cloud registration with a single Kinect camera was used to build this model, from which multi-view gait sequences were synthesised to extract view-invariant features. The analysis was done on a dataset of 100 subjects captured by a single Microsoft Kinect camera and a mean recognition rate of 89% was obtained over a range of synthesised views from 0° to 90°. Sivapalan et al. [106] proposed the Gait Energy Volume (GEV) by extending the concept of the GEI [34] to 3D space. This representation had been applied on both 3D gait data reconstructed from multiple cameras and partial volume reconstructions built from depth frontal images. The analysis on an in-house dataset of 15 subjects captured by the Microsoft Kinect showed a high discriminatory capability of the GEV where a recognition rate of 100% was achieved. Nakajima et al. [86] proposed a novel gait feature representation, called Depth-Based Gait Feature (DGF). In this work, depth images were aligned to cope with the problems of a view variation and perspective distortion. The position of a person in 3D space was firstly calculated and the depth image was then aligned accordingly. The new gait representation was finally computed by applying the Discrete Fourier Transformation to the aligned depth images in one gait cycle. Experimental results showed the effectiveness of this technique for person authentication. The main limitations of depth sensor-based approaches are that (1) they assume both gallery and probe are captured by a depth sensor, and (2) they cannot cope with large walking direction changes since a depth sensor records a person from his/her front.

2.5 Gait analysis techniques for curved trajectories

Recently, a few studies in the literature have been conducted in the field of gait recognition on curved trajectories. These studies have been carried out in parallel with the work done in this thesis. They used either 3D reconstructed data or multi-view gait sequences to tackle the problem of walking direction changes. Castro et al. [15] used motion descriptors to build a pyramidal representation of the gait motion. The person region in each frame was divided into several areas to extract the local motion descriptors, which were then combined into a final gait descriptor for recognition. Experimental results on the AVAMVG multi-view gait dataset showed that the technique achieved an average recognition rate of up to 91.6% for testing on curved trajectories and training on straight trajectories using a multi-camera setup and a majority voting strategy. López-Fernández et al. [76] introduced a rotation invariant gait descriptor based on analysing the temporal variation of 3D angular movement of the walking subject on multi-camera setup. A sliding temporal window for majority voting policy was used to smooth the classification results. An average recognition rate of 98% was obtained on the KY4D gait dataset for curved walks when using both straight and curved walks for training. However, the performance deteriorated for tilted cameras. Lpez-Fernndez et al. [77] introduced a multi-view gait recognition technique based on 3D reconstructions of people walking on unconstrained paths. In this technique, 3D models of people were aligned along their way and a new gait descriptor, called gait entropy volume [see Fig. 2.11], was extracted by exploiting the theory of entropy. Recognition results were computed using a Support Vector Machine and a sliding window for majority voting. A recognition rate up to 98% and 71% was obtained on the AVAMVG and KY4D gait datasets respectively for matching straight with curved walking. López-Fernández et al. [73] developed appearance-based gait descriptors based on 3D gait volumes which were previously aligned to account for walking direction changes. These descriptors were computed by aggregating all the cubes that can fit into a volume. An average recognition rate of 96.1% and 93.8% for matching straight with straight walking on the AVAMVG and KY4D gait datasets respectively while these numbers decreased to 63.2% and 71% for matching straight with curved walking. A recent study in [74] calculated a new rotation invariant gait descriptor which is based on dividing each 3D volume into a number of horizontal slices and computing the centroid of each slice. Then, a descriptor is defined as a tuple of the acute angles between the normal vector to the floor plane and the vector joining each pair of consecutive centroids. With these descriptors, an average recognition rate of 99.7% and 99.5% is recorded for matching straight versus straight walking on the AVAMVG and KY4D gait datasets respectively while a recognition rate of 93.6% and 72.8% is obtained for matching straight versus curved walking on these datasets.

Although the techniques in [15, 76, 77, 73, 74] were applicable for gait recognition on curved paths, however they required a multi-camera setup for both training and testing. Iwashita et al. [44] proposed a recursive image synthesised method by fitting a curve to the estimated foot position in each frame along one gait cycle. This method utilised 3D volumes of people walking along straight lines for training and 2D images captured by a single camera for testing. A synthetic image was generated from each 3D volume according to the estimated walking direction at each frame. The subjects were then identified using Affine Moment Invariants extracted from each image as gait features. An average recognition rate of 98% was achieved for matching straight with straight walking and 66.6% for straight with curved walking on the KY4D gait dataset. Although this method allows walking direction changes in one gait cycle, it appears to be computationally expensive and had not been used for recognising arbitrary view(s). This study will be considered for a comparison purpose in chapters 6 and 7 since it used similar problem settings to those adopted in this thesis.

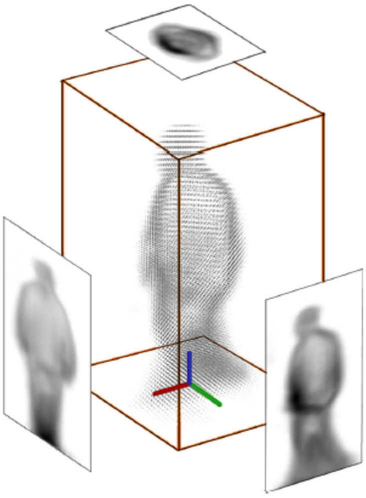


Figure 2.11: Gait Energy Volume and its projection from 3 orthogonal views in [77].

2.6 Discussion

This chapter describes current progress, a research background and various approaches for gait recognition under view variation. There is a considerable progress in the size of the collected gait datasets from tens to ten thousands of people. The current datasets also include a wide range of covariate conditions, which are essential for accurately estimating intra- and inter-class variation. Publicly available gait datasets are a fundamental requirement for benchmarking different techniques.

The discriminatory power of holistic-based features has shown to be better than that of model-based features as the former recorded higher recognition rates than the latter under similar conditions. Furthermore, holistic-based techniques directly extract gait features from the shape and motion information of the walking subjects, whereas model-based techniques extract features from the parameters of the fitted model. Gait features derived by holistic-based techniques are easy to compute whereas model-based techniques require higher resolution images and computational resources for better model fitting and complex searching and matching. However, they are less sensitive than holistic-based techniques to appearance variations due to small view changes.

The majority of gait recognition techniques recorded high recognition rates when comparing gait features under the same view. However, people in the real world can be captured from any camera viewpoint and their walking direction can also change. These factors adversely affect the performance of gait recognition techniques that are designed for a specific view [58]. Many approaches have been developed to cope with view variation. Some of these approaches build view-invariant features for cross-view matching. Although these techniques assume that gallery and probe samples are from different views, they perform well only for a limited range of view variations. Other techniques use a training process to learn the relationships of gaits observed across views. These techniques can cope with a larger view variation compared to view-invariant techniques. However, their performance deteriorated significantly when a view difference is greater than 30° [56]. A few approaches reconstruct a 3D human model from multiple synchronous cameras. Working in 3D domain overcomes the problems of view dependency and self-occlusion since gait under any view can be synthesised from the 3D model.

Using a multi-camera setup to build 3D models (volumetric data) in both enrolment and recognition phases is not a practical solution. Alternatively, people can be enrolled using their 3D models while matching can then be performed against gait images from single cameras using synthetic images derived from the 3D models. Only one study in the literature evaluated the performance of matching 3D against 2D gait data. This type of matching is highly suitable for forensic applications where matching needs to be computed for gait images from a surveillance camera under an arbitrary view.

Furthermore, most of the existing approaches and gait datasets assume that people walk only along straight lines. This is, however, the simplest style of walking. In reality, people change their walking direction from time to time to reach their destinations. When a person walks on a curved path, the observation angle in each frame is continuously changed which causes a gradual change in subject appearance. Little research has been carried out for gait recognition on curved paths. This may be due to the lack of a suitable gait dataset. Only two gait datasets include people walking on curved paths (the AVA Multi-View Dataset for Gait Recognition (AVAMVG) [75] and Kyushu University 4D gait Database (KY4D) [44]) are publicly available. The AVAMVG dataset contains a small number of people captured by only six cameras. Furthermore, the body parts of people, in this dataset, are partially seen by two cameras most of the time more than the remaining cameras [15]. On the other hand, the KY4D dataset includes more cameras and a larger number of people compared to the AVAMVG dataset. Therefore, the KY4D gait dataset will be used for the analyses in chapters 6 and 7.

Chapter 3

The Framework for Matching 3D with 2D Gait Data

According to the literature review in chapter 2, we observed that holistic-based features exhibit higher recognition performance than model-based features under similar conditions (i.e. no significant changes in the appearance of the subject). Therefore, holistic-based features are considered in this thesis for extracting gait patterns of the walking subjects. However, the appearance of the subjects undergoes variations when viewpoints and/or walking directions are changed (view variation), which alter the extracted features and ultimately reduce the recognition performance.

To deal with view variation problems, some techniques have been proposed by either extracting view-invariant features for cross-view matching or learning a mapping relationship between gaits under different views. However, these approaches perform efficiently for a limited range of applicable views. A few approaches reconstruct 3D volumetric data from multiple views to be used for view-invariant gait recognition since a gait image from any viewpoint can be synthesised from the 3D data. The main drawback of these approaches is that they require a cooperative multi-camera system which cannot always be available. Using such a system in forensic and security applications would be impractical. Instead, a multi-camera setup could be used during the enrolment phase only to register people in 3D. Recognition can then be done against gait images from a single camera (3D against 2D matching). Matching 3D against 2D gait sequences is used to solve the problem of view variation in this thesis.

This chapter describes the framework of matching 3D volumetric data with 2D silhouettes using the Biometrics Tunnel at Southampton University as a recording site. To evaluate the performance of matching 3D against 2D gait sequences, we adopted the Gait Energy Image (GEI) [72] as a gait feature due to its effectiveness, insensitivity to noise and wide use in the literature. Building a 3D sequence depends on extracting silhouettes from multi-view gait images and estimating calibration information of the recording cameras. These two operations are explained in the first part of this chapter. In the second part, we focus on analysing (1) the effect of different camera configurations on the 3D reconstruction process, (2) the discriminatory power of the GEI using silhouettes from different camera viewpoints and (3) the influence of using synthetic silhouettes. The performance of matching 3D volumetric data against 2D silhouettes from different viewpoints is discussed in the last part.

3.1 The original Biometrics Tunnel layout

The Biometrics Tunnel was established in an indoor laboratory at Southampton University for the purpose of automatic capturing video of a subject from multiple views. The Tunnel consists of a narrow pathway in the middle area, which is surrounded by two walls. The walls are painted with non-repeating patterns to facilitate the process of camera calibration. Three saturated colours have been chosen to simplify the separation of a subject from his/her background. Two infrared break-beam sensors were placed at the entry and exit points of the Tunnel to control the start and end of the data capture process. The first sensor is triggered when the subject enters the measuring area, and the acquisition process is started. the second sensors is triggered when the subject reaches the end of the measuring area, and is stopped capture the data. A large multi-coloured LED was placed at the entrance of the Tunnel to give an indication when the Tunnel is busy. The initial prototype of the Tunnel was constructed by Middleton et al. [83]. Initially, the Tunnel used nine cameras to capture video footage: eight synchronous cameras to record gait data and one camera for face and upper body. The gait cameras were distributed along the top of the two walls to provide good coverage for the Tunnel area. The frames from the gait cameras were streamed unprocessed to four computers via an IEEE1394 bus. There were also four synchronisation units to ensure simultaneous capture of the frames from all the cameras during the capture process.

The capturing system in the Tunnel was later altered and its layout was expanded by Seely [98] to result in a system capable of acquiring multi-biometrics data in a fast and efficient way. Four new gait cameras were added at the far bottom corners of the Tunnel to improve the accuracy of the 3D reconstruction process and one additional camera was also used to capture the imagery of a subject's ear. The gait cameras were connected to a single hub in threes. In total, there were four computers to store the gait images from the twelve cameras. In addition, a more powerful computer was used to manage the capturing process of the Tunnel system and perform the 3D reconstruction. All the gait cameras have a resolution of 640×480 pixels with a rate of thirty frames per second. The four gait cameras in the middle of the Tunnel have a wide-angle lens which exhibits a higher amount of distortion, while the remaining gait cameras in the far-end have a small-angle lens and a lower amount of distortion. Fig. 3.1 shows the arrangement of the cameras inside the Tunnel.

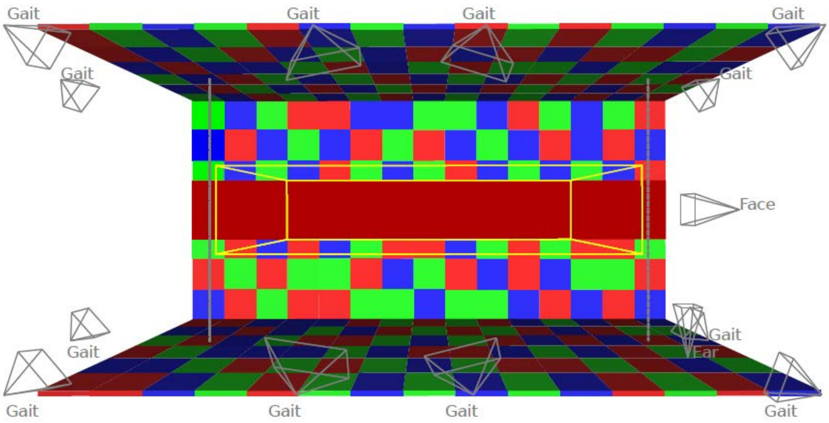


Figure 3.1: The placement of the cameras in the Biometrics Tunnel [98].

3.2 Matching stages of 3D with 2D gait data

Fig. 3.2 shows the basic stages for recognising 2D silhouettes from a single camera(s) using 3D volumetric data. During the enrolment stage, the synchronous cameras are used to record peoples' gait images from different angles. The silhouettes are extracted from these images and used to build 3D models of people. One gait cycle is then determined from each 3D walking sequence as will be explained in section 3.2.2. Next, a synthetic silhouette is generated from each 3D model (volume) along one gait cycle using a probe camera projection matrix (i.e. camera parameters). After that, gait features are extracted from the synthetic silhouettes and used to build a gallery set. In the testing phase, subject silhouettes are captured from an arbitrary camera, and one gait cycle is determined from each sequence. Gait features are then extracted from the real silhouettes and used to build a probe set. During the recognition phase, gait features from the gallery and probe sets are matched based on a minimum distance, and the recognition results are computed using nearest neighbours classifier. The details of the matching process will be explained in the following subsections.

3.2.1 3D Reconstruction process

To build a 3D volume of a subject, twelve synchronous cameras are used to record a subject's images from different viewing angles. Silhouette images are then extracted using a background segmentation algorithm. Calibration information of the synchronous cameras is also calculated in order to back-project each silhouette into 3D space. Finally, the intersection of the back-projected silhouettes forms a 3D volume of the subject [62]. Fig. 3.3 shows a diagram of the 3D reconstruction process. The detailed description about the basic steps of the 3D reconstruction process will be explained in the following subsections.

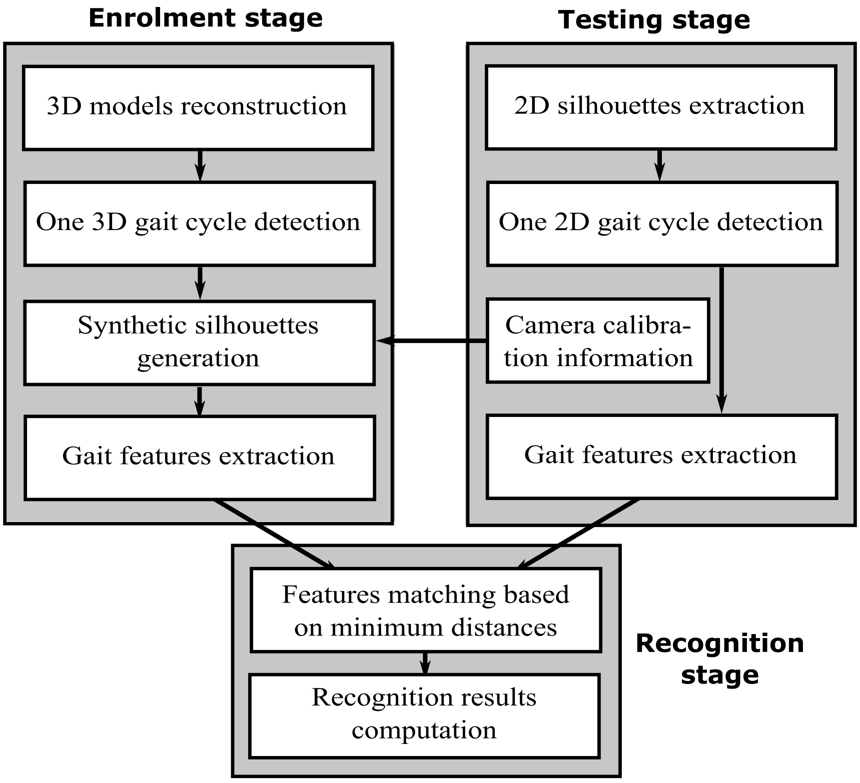

Enrolment stage
3D models reconstruction
One 3D gait cycle detection
Synthetic silhouettes generation
Gait features extraction
Testing stage
2D silhouettes extraction
One 2D gait cycle detection
Camera calibration information
Gait features extraction
Features matching based on minimum distances
Recognition results computation
Recognition stage

Figure 3.2: Flowchart of matching 3D with 2D gait data.

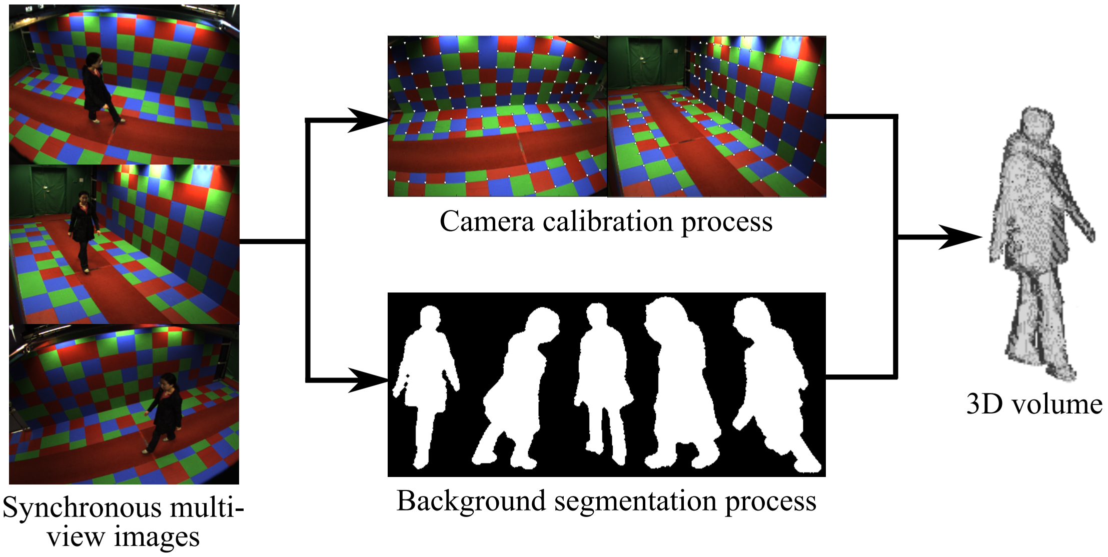

Synchronous multi-view images
Camera calibration process
Background segmentation process
3D volume

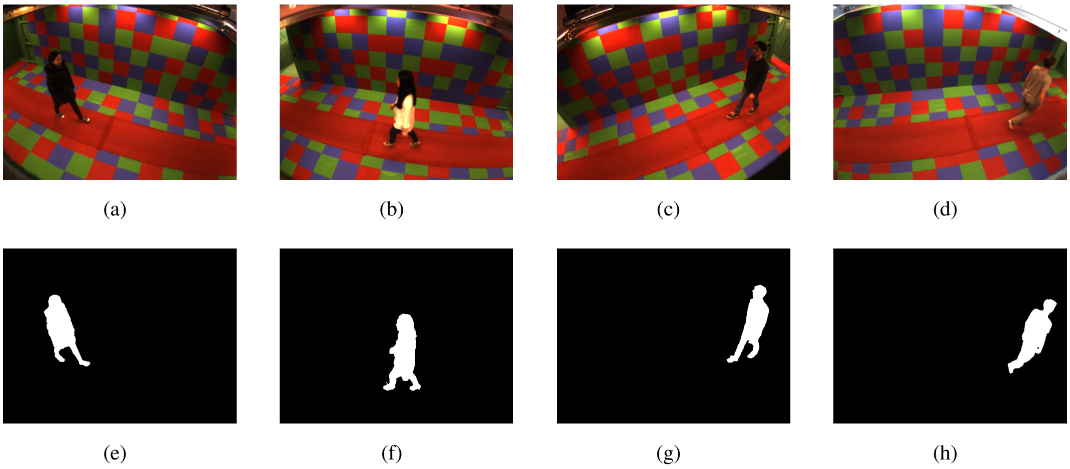
Figure 3.3: Basic steps in the 3D reconstruction process.

3.2.1.1 Background segmentation

Background segmentation is the first step in any holistic-based gait recognition technique. Given a sequence of images captured by a static camera, the silhouette images can be generated through a background segmentation, which consists of background modelling and subtraction processes. Since the Biometric Tunnel represents a constrained environment where lighting is controlled, the background is static and camera settings and parameters are fixed, no sophisticated background segmentation algorithm is required. First, a partially-normalised colour space (RGB) is calculated as

$$\acute{R} = \frac{R}{0.333 * (R + G + B)}, \quad \acute{G} = \frac{G}{0.333 * (R + G + B)}, \quad \acute{B} = \frac{B}{0.333 * (R + G + B)} \tag{3.1}$$

The purpose of using the partially-normalised colour space is to reduce the effect of shadow in the segmented image. Then, a proper estimation for background modelling is statistically built by computing the distribution of pixel intensities over time according to Gaussian distribution. This can be done by recording a number of empty frames and computing the mean value (μ) for these frames over each normalised colour channel as a background model. After that, the pixel in the current frame ($F(t)$) is labelled as a foreground by comparing the distance between its value in the current frame and the background model against a predetermined threshold. However, the selection of a suitable threshold value is critical, especially in a case of a low-contrast image. Using a single threshold value for all pixels may introduce some background segmentation errors as it does not consider how noise properties for the pixels across the sensor are varied. Instead, the segmentation threshold value is chosen on a pixel by pixel basis by computing the standard deviation (σ^2) for each background's pixel and multiplying this number by a constant, k, to produce a suitable threshold value for each pixel. The segmented silhouette image is computed as $||F(t) - \mu|| > k \, ||\sigma^2||$ where $||X||$ =$\sqrt{X_R^2 + X_G^2 + X_B^2}$ (i.e. is the square root of the sum of square values in the R, G, and B components of the signal X) and $k = 1.2$. The value of k is chosen according to the quality of background segmentation results for several sequences. Binary dilation followed by binary erosion with a structuring element of size 5×5 pixels are then used to clean up the artefacts from the segmented silhouette image. Finally, a binary connected component analysis is applied to extract the largest connected region as a silhouette. Fig. 3.4 shows examples of background segmentation results for different cameras in the Tunnel.


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

(e) (f) (g) (h)

Figure 3.4: Background segmentation results. First row (a-d) includes foreground colour images and second row (e-h) shows their corresponding silhouette images.

3.2.1.2 Camera calibration

The main aim of camera calibration is to determine its position, orientation and internal properties that are required to do the mapping from a point in a 3D scene to a corresponding point in a 2D image. To compute the calibration information of a camera, a pinhole model [107] is often used to approximate the camera. The pinhole model is simple but precise enough to be used in many applications. In this model, the camera consists of a black box with a hole through which rays of light from the outside world pass and are projected onto the image plane. This hole is called the camera centre, which represents a centre of projection.

The pinhole model assumes that the image axes are aligned with the X_c- and Y_c-axis of the camera coordinate system. The optical axis of the camera is aligned with its Z_c-axis as shown in Fig. 3.5. To compute a mapping of a point from the world coordinate system to the image plane, the translation and rotation are used to put the point in the camera coordinate system before projecting it onto the image plane using perspective projection transform [107, 36]. The translation and rotation are computed using 3×1 translation vector (t) and 3×3 rotation matrix (R) respectively. The parameters containing in the R and t are called camera extrinsic parameters and are used to align the origin and axes of the world coordinate system to that of the camera coordinate system. The R and t are combined into a 3×4 transformation matrix as E = [R | t].

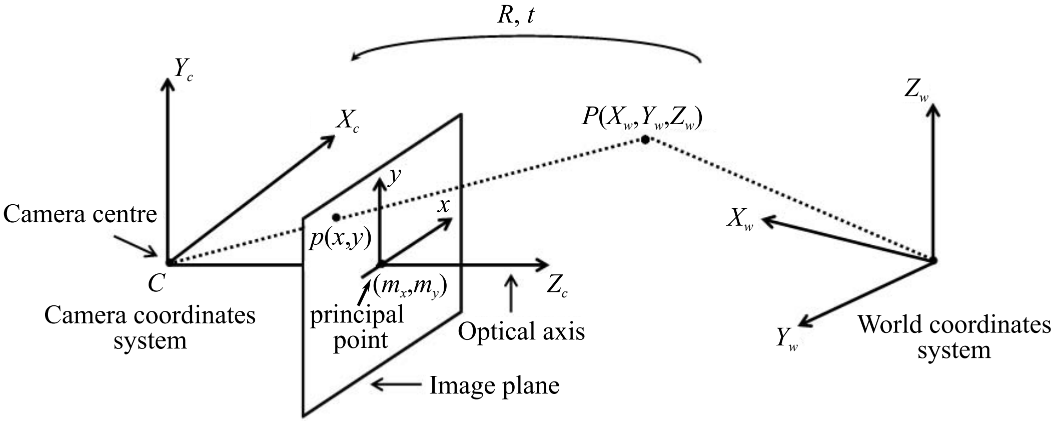


Figure 3.5: Perspective projection of a point in a 3D world coordinates onto a corresponding point in a 2D image plane.

To project the point onto the image plane using perspective projection transform, the factors f_x and f_y are used to change the measurements made in the camera coordinate system to the equivalent one used in the image coordinate system (i.e. in pixels) where $f_x = f\ \alpha_x$ and $f_y = f\ \alpha_y$ are the focal length of the camera in terms of pixel dimensions in the x and y directions respectively. The f represents the distance (in mm) between the camera centre and the image plane while α_x and α_y represent the number of pixels per unit distance in image coordinates along the x and y directions respectively. Most the current imaging system define the origin of the image coordinate system at the left-top pixel in the image. However, in the pinhole model it is assumed that the origin is at the principal point (m_x, m_y) in which the optical axis intersects the image plane as shown in Fig. 3.5. Therefore, the perspective projection transform can be represented using the following 3×3 matrix

$$A = \begin{bmatrix} f_x & s & m_x \\ 0 & f_y & m_y \\ 0 & 0 & 1 \end{bmatrix} \tag{3.2}$$

where s represents the skew factor. Most digital cameras have square pixels in the camera sensor and no skewing. The above matrix (A) called the camera calibration matrix and is used to describe the intrinsic camera parameters. Using the E and A matrices, the final camera projection matrix can be described as a 3×4 matrix, $M = A\,[R \mid t]$. The point P in the world coordinate system is projected to the image point p (both expressed in homogeneous coordinates) using the projection matrix M as [36]

$$\lambda p = MP \tag{3.3}$$

The scalar λ is the inverse depth of the 3D point. It is required when all coordinates are homogeneous [1] and the last element is normalised to one.

The previous model for the pinhole camera is ideal and is based on the constraint that the straight lines in 3D scene must be projected as straight lines in the 2D image plane. Unfortunately, most of the available lenses in the markets introduce some form of a deviation that makes these straight lines appear as curves in projected image planes [14, 26]. Generally, there are two types of lens distortion: radial and tangential. Fig. 3.6 shows examples of the radial lens distortion. Radial distortion is a deviation along the direction from a centre of distortion (or projection) to the intended image point, whereas tangential distortion is a deviation perpendicular to that direction.

For a camera with a highly distorted (low-cost or wide-angle) lens, both radial and tangential distortions can exist [26]. In order to achieve accurate calibration results, we took these two types of distortions into consideration. The radial and tangential distortion are modelled and their coefficients are estimated for all cameras in the Tunnel.

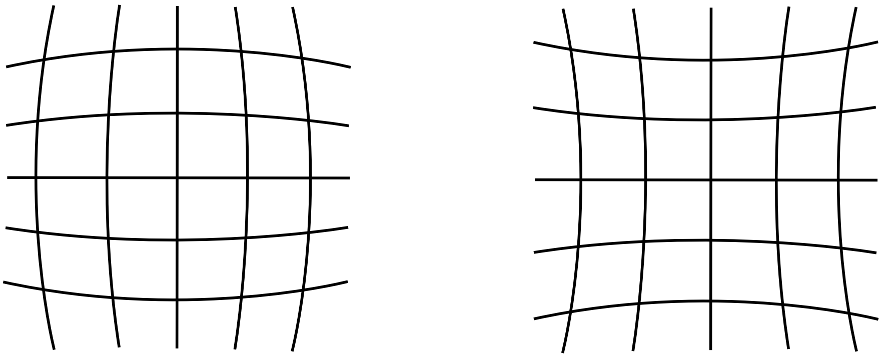
The radial distortion is modelled using the following nonlinear transformation [14]:

$$\begin{aligned} x_u &= x_d(1 + k_1 r_d^2 + k_2 r_d^4 + k_3 r_d^6 + ...) \\ y_u &= y_d(1 + k_1 r_d^2 + k_2 r_d^4 + k_3 r_d^6 + ...) \end{aligned} \tag{3.4}$$

Where (x_d,y_d) and (x_u,y_u) are the image coordinates in the distorted and undistorted (corrected) images respectively. k_1 , k_2 and k_3 are the radial distortion coefficients, and r_d represents the distance between the distortion centre and a point (x_d,y_d) in the distorted image, which can be described as Euclidean distance:

$$r_d = \sqrt{x_d^2 + y_d^2} \tag{3.5}$$

[1] The homogeneous coordinate is achieved by adding an extra element to the original coordinate.



(a) Barrel distortion

(b) Pincushion distortion

Figure 3.6: Examples of radial lens distortion.

The tangential distortion, on the other hand, is modelled mathematically as [26]:

$$\begin{aligned} \delta_x &= p_1(3x_d^2 + y_d^2) + 2p_2x_dy_d + \cdots \\ \delta_y &= 2p_1x_dy_d + p_2(x_d^2 + 3y_d^2) + \cdots \end{aligned} \tag{3.6}$$

Where δ_x and δ_y are the horizontal and vertical tangential distortion components respectively. p_1 and p_2 are the tangential distortion coefficients. Radial and tangential distortions can be combined together using the following pair of equations:

$$\begin{aligned} x_u &= x_d + k_1x_dr_d^2 + k_2x_dr_d^4 + k_3x_dr_d^6 + \cdots + p_1(3x_d^2 + y_d^2) + 2p_2x_dy_d + \cdots \\ y_u &= y_d + k_1y_dr_d^2 + k_2y_dr_d^4 + k_3y_dr_d^6 + \cdots + 2p_1x_dy_d + p_2(x_d^2 + 3y_d^2) + \cdots \end{aligned} \tag{3.7}$$

In this thesis, we considered three coefficients for radial distortion and two coefficients for tangential distortion to model camera distortion using Eq. 3.7. We also assumed the projection centre to be at the principal point. In order to derive the values of the projection matrix and lens distortion coefficients for each camera, a calibration grid with well-known points' positions is required to extract a number of 2D-3D corresponding points. The Biometric Tunnel is used as a calibration grid, where the walls and floor are painted with non-repeated patterns to facilitate the computation of the calibration parameters for each camera. Fig. 3.7 shows the global (reference) coordinate system used in the Tunnel to extract the corresponding 2D-3D points.

Using the 2D-3D corresponding points that are determined manually, a set of horizontal and vertical lines in a distorted image is defined. Then, the distance between the line joining its two end points in the image and its mid-point is used as an optimisation function to minimise the curvature of the lines in the distorted image. The optimisation is performed using the Nelder-Mead simplex optimisation algorithm [87], which is implemented using SciPy package [50] in Python language. The results of the optimisation are used to find the initial values for the lens distortion coefficients and straighten the curved lines according to Eq. 3.7. Fig. 3.8 (a) shows an

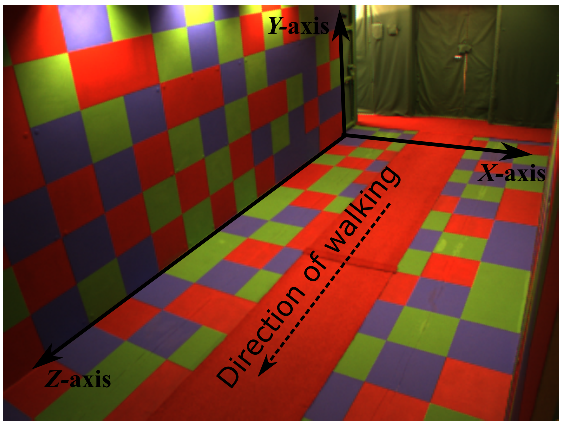


Figure 3.7: The global 3D coordinate system in the Tunnel.

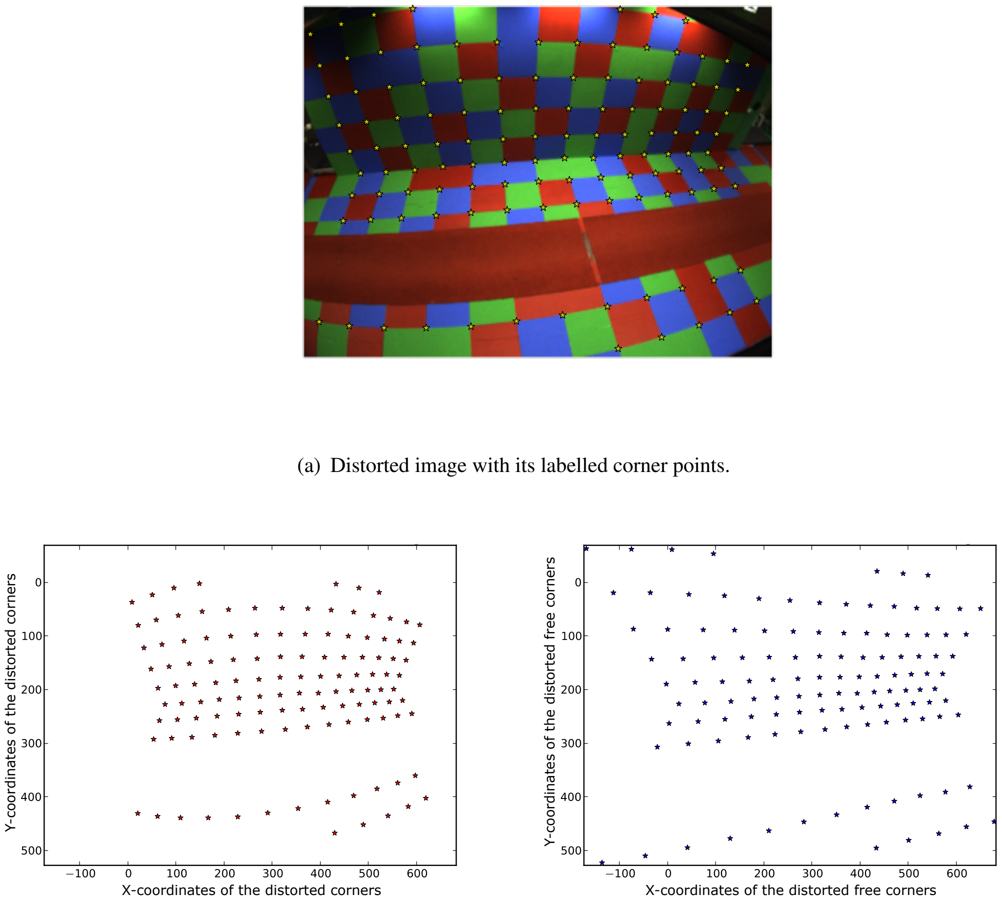
example of a highly distorted image captured by one of the middle cameras in the Tunnel, where a set of corner points is manually labelled. Fig. 3.8 (b) and (c) display the positions of these points in the distorted and corrected images respectively. From Fig. 3.8, it can be seen that the positions of the corner points lie on the straight lines after removing the effect of distortion.

After that, a set of the resulting 2D points in the corrected image with their corresponding 3D points was used to find the initial values of the projection matrix M by establishing a number of equations using Eq. 3.3 and solving these equations using SVD. These values are optimised by minimising the mean of the total sum of square distances between the projected 3D points onto the image plane and the corrected 2D points. Finally, the distortion coefficients are considered together with the projection matrix to iteratively minimise the mean error. As explained, the optimisation process was performed hierarchically and iteratively due to the higher number of unknown variables. In such case, a direct optimisation would be unreliable.

3.2.1.3 Building a 3D volume from multi-view silhouettes

A 3D volume of a human body can be reconstructed using three or more synchronous silhouettes captured from different viewpoints. The silhouette from each view is back-projected into 3D space using camera calibration information to produce a 'cone'. The intersection of the cones from all views form a visual hull [62], which approximates the 3D volume of the human. The resulting volume is divided into a 3D grid of equally cubic elements called voxels. Each of these voxels is then tested for its occupancy in a visual hull region as either occupied, partially-occupied, or empty [16]. This technique is known as shape from silhouette reconstruction. In our implementation, the voxel is marked as occupied if all the twelve cameras in the Tunnel see the foreground pixel at the corresponding positions in all images.

In the Biometric Tunnel, the shape from silhouette approach is used for 3D reconstruction due to its flexibility and simplicity, although the direct implementation of such approach is very costly because it requires repeatedly computing the mapping from 3D world coordinates to 2D image coordinates for each camera. To improve the processing speed, the reconstruction process can



(a) Distorted image with its labelled corner points.

(b) Corner points' positions in the distorted image.

(c) Corner points' positions in the lens corrected image.

Figure 3.8: Distortion correction.

be done using lookup tables. Assuming that there is no change in the cameras' positions and orientations, the lookup table is first established for each camera to pre-compute the mapping from 3D world coordinates to 2D image coordinates where the entries to the lookup table are the positions in 3D world coordinates while its contents are the the positions in image coordinates. The intersection of back-projected silhouettes can then be computed as shown in Fig. 3.9. The algorithm that carries out the 3D reconstruction process using lookup tables is illustrated in Appendix A. An example of a 3D volume reconstructed using the intersection of the cones from all the 12 cameras in the Tunnel is shown in Fig. 3.10. The time required to reconstruct one volume of size 240 × 286 × 600 using lookup tables is about $1.3sec$ using all the 12 cameras in the Tunnel. The size of each voxel is $1cm^3$.

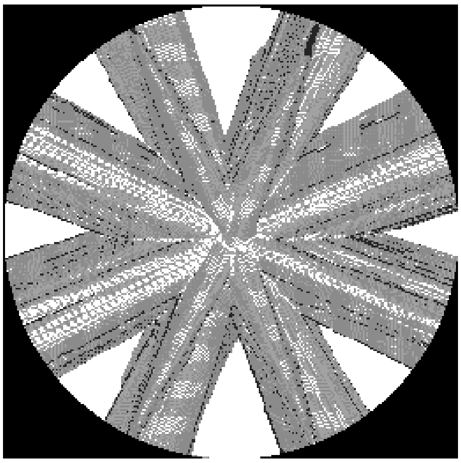


Figure 3.9: Intersection of cones from several viewpoints (top view).

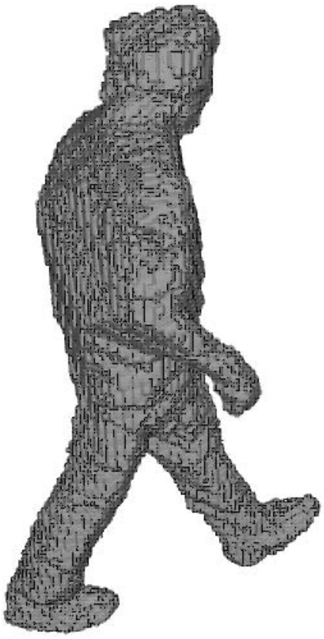


Figure 3.10: Example of a 3D volume reconstructed using the 12 cameras in the Tunnel.

3.2.2 3D Gait cycle estimation

Human walking can be described as a sequence of limbs' motion that is done in a repeatable and distinctive way for each individual [54]. Therefore, a gait cycle can be extracted from a gait sequence to describe a human's walking pattern and the gait features can then be computed from the frames within a gait cycle to recognise humans by their gaits. The gait cycle serves to align each gait sequence before matching. One of the methods used to determine a gait cycle is to analyse the variation of the width of a bounding box. These variations can contain structural and dynamic information about the gait. The bounding box width reaches a maximum when the two legs are farthest apart (double support stance) and drops to a minimum when the legs overlap (mid-stance). For a 3D volumetric sequence, we analysed the variation in the width of a 3D bounding box around the lower body region to detect the first and last frame in a gait cycle where the width of the bounding box is parallel to the direction of walking in the Tunnel as shown in Fig. 3.7.

Fig. 3.11 shows the variation of a typical bounding box width along the direction of walking for a 3D sequence, where the bounding box is determined for the lower region (17% of height) of each 3D volume. Several peaks and valleys can be detected in this figure that correspond to the phases when the two legs extend to a maximum and contract to a minimum respectively. A Gaussian filter with sigma 1.5 is used to remove the outliers and smooth out the curve. A complete gait cycle is then labelled using three consecutive local minima (i.e. mid-stance phases) [see Fig. 3.12] of the bounding box width curve because these are more stable than the local maxima. The selection of the start and end of a gait cycle when the two legs completely overlap is required in the latest work in this thesis. More than one gait cycle can be detected in each 3D sequence. However, one gait cycle closest to the centre of the Tunnel is selected from each 3D sequence to ensure the chosen cycle includes a complete 3D volume in all frames. To determine the closest gait cycle to the centre of the Tunnel, the centroid of the middle volume in each gait cycle is extracted and the distance between the centroid of the volume and the central point (143,120,300) in the Tunnel is calculated. The gait cycle with the smallest distance to the central point is selected as the closest gait cycle to the centre of the Tunnel.

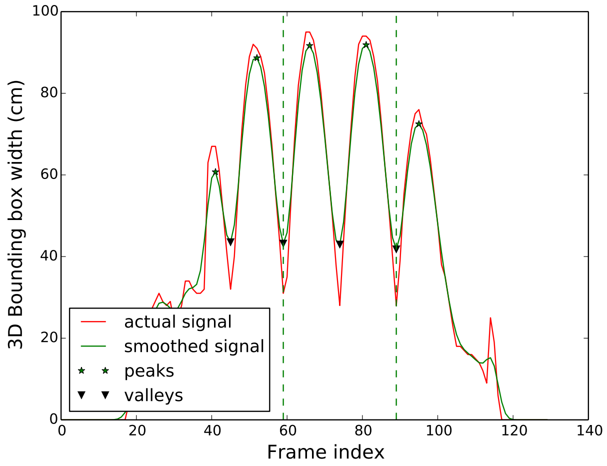


Figure 3.11: The variation of bounding box width of the lower limbs. The dashed lines represent the start and end of one gait cycle.

3.2.3 Gait features extraction

The Gait Energy Image (GEI) [72] is used to extract a gait feature from one gait cycle in each sequence in the gallery and probe sets as it is effective, insensitive to noise in an individual silhouette, and widely used in the literature as a baseline gait feature for single-view and multi-view gait recognition. In a GEI, a single grayscale image is calculated by computing the average of all silhouette images over a complete gait cycle. Let $S(1)_{x,y}$ be the first silhouette in the cycle

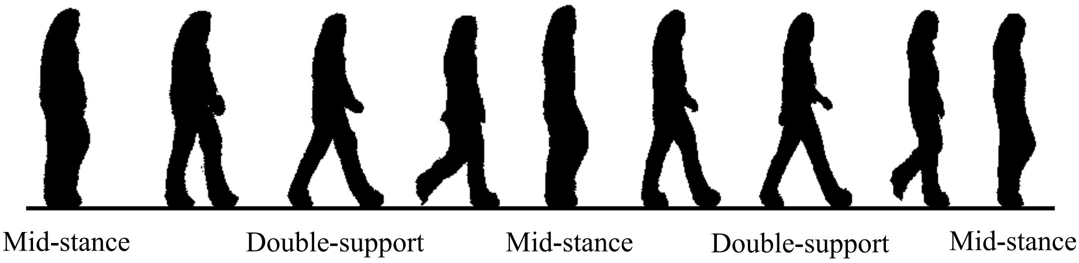


Figure 3.12: A complete gait cycle contains three mid-stances for a female wearing tight clothing.

and $S(L)_{x,y}$ be the last silhouette, the GEI can be calculated using

$$GEI = \frac{1}{L} \sum\nolimits_{t=1}^{L} S(t)_{x,y} \tag{3.8}$$

Where L is the number of silhouettes in one gait cycle. Two versions of GEIs are considered: normalized and non-normalized GEIs. The non-normalized GEI can be computed by centring each silhouette in one gait cycle such that the centroid of a silhouette is at the middle of the image. The resulting silhouette image is then cropped using a smallest bounding box that incorporates the whole silhouette region. Finally, the average image is computed from all the cropped silhouettes in one gait cycle. This representation has the advantage of preserving subject's height. On the other hand, the normalized GEI is computed by normalizing the size of each silhouette such that all the silhouettes have the same height by keeping the aspect ratio of the width-to-height constant throughout the gait cycle. For example, if the size of the silhouette is resized by a factor of 2 along the vertical direction, then the same scaling factor is also applied along the horizontal direction. The non-normalized GEI is used in this thesis only with the orthogonal side-view synthetic silhouettes which are generated from the 3D volumetric sequences because the height of these synthetic silhouettes is constant regardless of their distance from the camera as will be discussed in section 3.3.1. Fig. 3.13 shows an example of a normalized GEI computed from silhouettes under two different view. As can be seen, the higher intensity pixel values in the GEI correspond to the body parts that undergo little relative movement (e.g. torso and head); whereas the dynamic body areas which have relative motions during walking (e.g. lower parts of limbs) have smaller values. The GEI thus captures both global body shape and motion information. For a comparison purpose, all the GEIs are rescaled to 64 × 64, resulting in 4096 features for classification.

3.3 Recognition results

All the experiments are carried out using the Soton multi-view gait dataset [98]. A small subset containing 43 subjects was selected for our analysis because the videos were all captured within a small time period, to ensure that the calibration of the cameras did not change significantly. These subjects were recorded using the twelve synchronous cameras in the Tunnel, and the

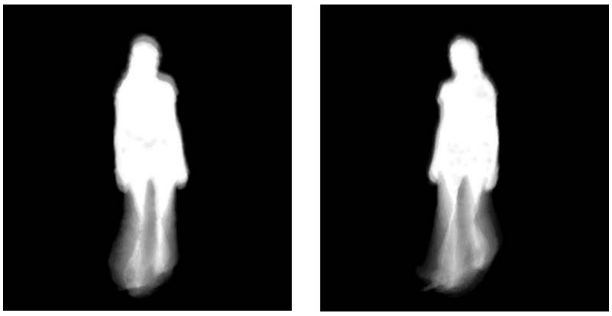


Figure 3.13: Examples of GEIs from two different camera viewpoints.

recording was done in several sessions. Six sequences were chosen manually for each subject after excluding the sequences with badly segmented silhouettes. The selected dataset contains 258 multi-view sequences for analysis.

The recognition results are evaluated using the nearest neighbours classifier. We used a rule similar to the leave-one-out recognition. For each sequence in the probe set, we removed the corresponding sequence in the gallery set that was captured at the same time. In order to compute the nearest neighbours classification, a distance matrix was calculated by computing the Euclidean distances between all pairwise GEIs in the gallery and probe sets. Let GEI_p^i and GEI_g^j are the GEIs at index i and j in the probe and gallery sets respectively, the Euclidean distance $(dist)$ is calculated as

$$dist(GEI_p^i, GEI_g^j) = ||GEI_p^i - GEI_g^j|| = \sqrt{\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (GEI_p^i(m,n) - GEI_g^j(m,n))^2} \quad (3.9)$$

where (m, n) is a specific pixel location in the GEI image while $M \times N$ are the dimensions of the GEI image. To identify unknown GEI in the probe set, the Euclidean distances between this GEI and each of the GEIs in the gallery set are measured. Then, the best match in the gallery set is defined as the one with the smallest Euclidean distance. The Non-normalized GEI is used only for analysing camera configurations effect as will explained in section 3.3.1; whilst the normalized GEI is used for the other analyses in this chapter due to the variation in subject height with respect to the distance from the camera. The results are evaluated using Correct Classification Rate (CCR), Receiver Operating Characteristic (ROC) curve, Equal Error Rate (EER) and decidability index $(d^{'})$.

The CCR is a common metric to evaluate the performance of a recognition system . The values of the distance matrix are used to compute the CCR. The CCR is defined as the percentage of the correctly classified samples in the probe set. For an identification problem, some outcomes can be provided by the recognition analysis: true positive, false positive and false negative rate. True positive rate is the proportion of samples of probe subjects that are correctly matched with their enrolled samples in the gallery set. False positive rate is the proportion of samples of different subjects in the probe and gallery sets that are incorrectly matched. False negative rate is the proportion of samples of same subjects that fail to match. The ROC curve is a measure of the verification performance. The ROC curve shows a graphical relationship between false positive and true positive rate in the horizontal and vertical direction respectively by varying the value of a matching threshold. The faster the ROC curve approaches one, the better the matching algorithm. On the other hand, Equal Error Rate (EER) provides a fair measure for comparing the performance of various techniques [38]. This metric is defined as the rate at which false positive and false negative rate are equal. The other metric that is used to assess the recognition performance is the decidability index $(d^{'})$ [25]. To compute this metric, two regions in the distance matrix can be identified: intra-class and inter-class. The intra-class region represents the distance values between samples of the same subjects while the inter-class region represents the distance values between samples of different subjects. The mean and standard deviation for these regions are used to compute the decidability index, which measures the amount of overlap (or separation) between intra-class and inter-class distributions. The higher value of the decidability index the better the matching algorithm. The decidability index is computed as $(d^{'})$ [25]

$$d^{'} = \frac{|\mu_{intra} - \mu_{inter}|}{\sqrt{(\sigma^2_{intra} + \sigma^2_{inter})/2}} \tag{3.10}$$

where μ_{intra} and μ_{inter} are the mean value for intra-class and inter-class distributions respectively, while σ^2_{intra} and σ^2_{inter} are the standard deviation for intra-class and inter-class distributions. Intra/inter-class distribution (variation) diagrams are also generated for some results. Several experiments have been conducted to analyse the influence of different factors on the performance of matching 3D against 2D gait data. These factors include the influence of different camera configurations in the Tunnel on the 3D reconstruction quality, the variation of the GEI performance using real silhouettes from different camera viewpoints and the effect of using synthetic silhouettes on the GEI performance. Finally, the performance of matching 3D volumetric data against 2D silhouettes from different viewpoints is measured. The details of these analyses will be illustrated in the following subsections.

3.3.1 Analysis of cameras configurations effect

An experiment was conducted using four different camera configurations to predict the impact of the number and placement of the cameras on the recognition performance. The first configuration (A) includes all the twelve cameras; the second configuration (B) includes the eight cameras distribute along the top of the Tunnel; the third configuration (C) includes the eight cameras at the far-ends of the Tunnel; and the last one (D) includes the four middle and the four far-bottoms cameras. Fig. 3.14 shows the placement of the middle and far-ends cameras in the Biometric Tunnel while Table 3.1 shows which cameras in each configuration. The volumetric data were reconstructed from multi-view gait sequences for each of these configurations. For a comparison purpose, orthogonal side-view silhouettes SD were synthesised from the 3D volumetric data as

$$SD_i(Y,Z) = \begin{cases} 1, & \text{if } \sum_{X=x_{min}}^{x_{max}} V_i(X,Y,Z) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{3.11}$$

Where SD_i is the synthetic orthogonal side-view silhouette from a 3D volume V_i. The above formula means that the synthetic side-view can be calculated from the voxels along the X-axis in the volume V_i, where the X-axis spans from the left to the right, the Y-axis is the vertical axis, and the Z-axis traverses the depth of the Tunnel. This view is not seen by any camera and thus it is a good way of testing the comparison of different configurations.

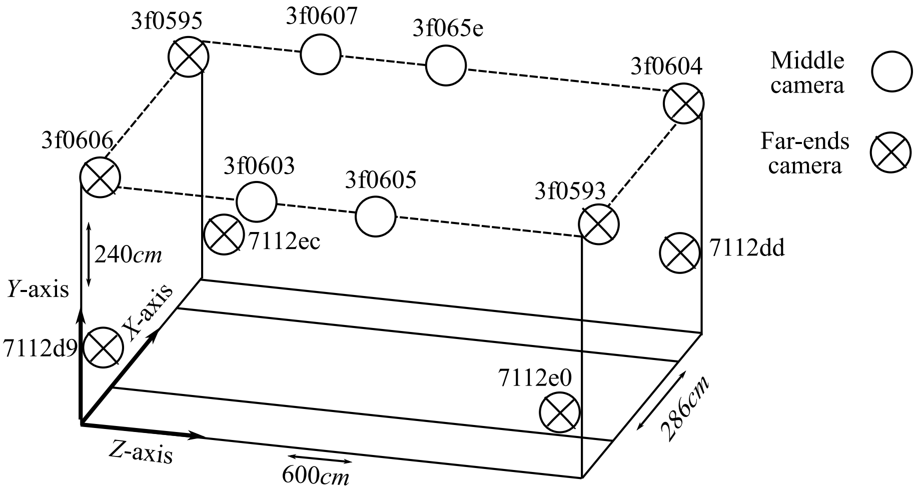


Figure 3.14: The placement of the middle and far-ends synchronous cameras in the Tunnel.

Table 3.1: Description of the camera configurations in the Tunnel.

Camera Configuration	Description	Camera ID
Configuration (A)	all 12 cameras	(3f0603, 3f0605, 3f0607, 3f065e, 3f0604, 3f0606, 3f0593, 3f0595, 7112d9, 7112dd, 7112e0, 7112ec)
Configuration (B)	8 top cameras	(3f0603, 3f0605, 3f0607, 3f065e, 3f0604, 3f0606, 3f0593, 3f0595)
Configuration (C)	8 far-ends cameras	(3f0604, 3f0606, 3f0593, 3f0595, 7112d9, 7112dd, 7112e0, 7112ec)
Configuration (D)	4 middle plus 4 far-bottom cameras	(3f0603, 3f0605, 3f0607, 3f065e, 7112d9, 7112dd, 7112e0, 7112ec)

A non-normalized GEI was computed from one gait cycle in each synthetic side-view gait sequence as the height of the subject using this viewpoint does not vary with respect to the distance from the camera. The recognition results for different camera configurations are shown in Table 3.2 and Fig. 3.15. As can be seen, all the configurations have similar recognition performance in terms of CCR. However, removing the four middle cameras has a positive impact on the recognition performance in terms of EER and ROC. It is expected that including more cameras would improve the quality of 3D reconstruction and the recognition results. Therefore, further investigation was carried out. We visually inspected several 3D volumes from each configuration. Fig. 3.17 shows that there is a huge amount of distortion in the 3D shape of a subject when discarding the four middle cameras whilst no noticeable distortion is observed when the other cameras in the Tunnel are omitted. It is hypothesised that the higher distortion in the 3D shape may increase the variation between the subjects, which improved the recognition performance. This is confirmed in Fig. 3.16, where the distributions of intra/inter-class variations are shown when including all 12 cameras and when discarding the middle cameras. It is observed from this figure that the distribution of the inter-class variation when removing the central cameras is wider than that observed when using all the twelve cameras.

The results obtained from this experiment show high recognition rates for the different camera configurations because the same type of data are used in both gallery and probe sets for matching. Furthermore, the highly distorted 3D shape obtained by excluding the middle cameras emphasises the importance of including these cameras for a better quality of the 3D volumetric data.

Table 3.2: Recognition performance using different camera configurations; using (A) all 12 cameras, (B) 8 top cameras, (C) without middle cameras, and (D) 4 middle and 4 far-bottom cameras.

Camera Configuration	CCR(%)	EER(%)	Decidability
12 cameras (A)	99.6	2.0	12.0
8 cameras (B)	99.6	2.1	11.9
8 cameras (C)	99.6	1.6	11.6
8 cameras (D)	99.6	2.2	12.0

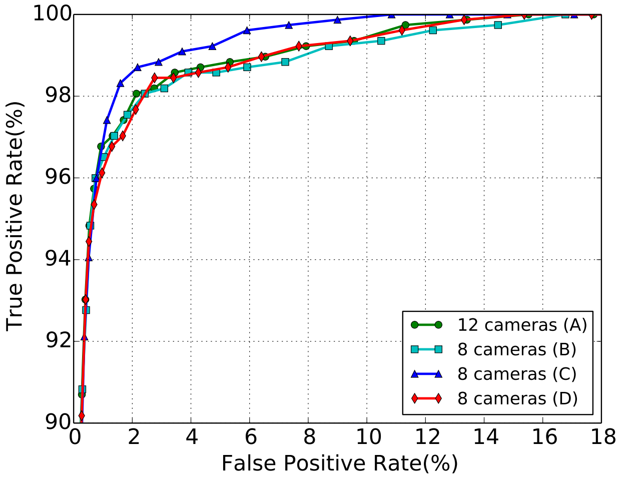
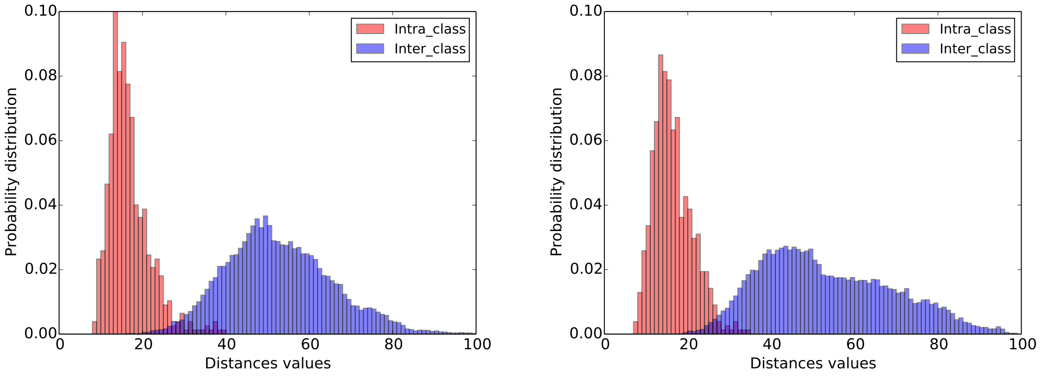


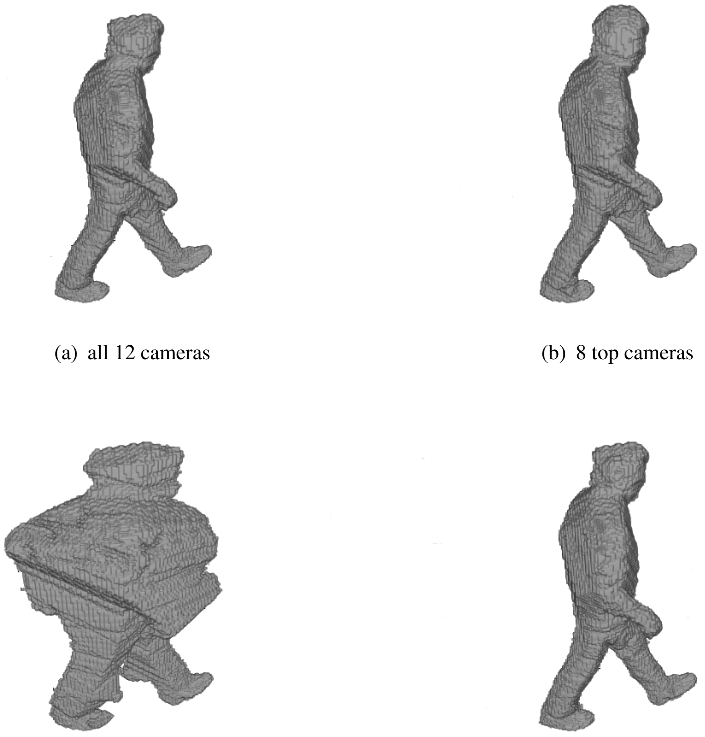
Figure 3.15: ROC curve for recognition performance using different camera configurations.



(a) 12 cameras

(b) 8 far-ends cameras

Figure 3.16: Intra/Inter-class distribution for two different configurations:(a) includes all 12 cameras; and (b) excluding the middle cameras.



(a) all 12 cameras

(b) 8 top cameras

(c) 8 far-ends cameras

(d) 4 central and 4 far-bottom cameras

Figure 3.17: 3D volume reconstruction using different camera configurations.

3.3.2 GEI performance using silhouettes from different viewpoints

There are two sets of cameras in the Tunnel as explained previously. The first set includes 4 wide-angle lens cameras in the middle of the Tunnel (close to the walking subject) and the second set involves 8 narrow-angle lens cameras at the far-ends of the Tunnel [see Fig. 3.14]. These sets help to evaluate the recognition performance of GEI using real silhouettes captured by cameras with different settings to assess the effectiveness of this gait representation for real world applications. An experiment was conducted in which the normalized GEI was computed for gait sequences from each camera in the Tunnel and the corresponding recognition performances were evaluated. The recognition results are shown in Table 3.3. As can be seen, the results split into two groups: the lowest recognition performance is recorded for the four central cameras in the middle of the Tunnel with the best recognition rate of 82.1%; whilst the highest performance is obtained from the far-ends cameras with the highest recognition rate of 100%. The ROC curves are shown in Fig. 3.18, which further confirm the results obtained in Table 3.3. The ROC curves for the middle cameras are significantly different from the ROC curves for the far-ends cameras. The type and position of the cameras cause a dramatic change in the shape of the ROC curve. To illustrate the difference in a class variation for the results obtained from the two sets of cameras, Fig. 3.19 shows an example of the class variation from one camera in each set. A smaller overlap can be seen between class distributions for the far-end camera as compared to that obtained from the middle camera.

These results reveal that the discriminatory power of the GEI is high for gait silhouettes from far and narrow field-of-view cameras. However, its discriminatory power significantly degrades for near and wide field-of-view cameras.

Table 3.3: Recognition results of the GEI for different camera viewpoints.

Camera ID	placement	CCR(%)	EER(%)	Decidability
7112e0	bottom-front-left	99.6	1.5	11.8
7112dd	bottom-front-right	99.6	1.7	11.4
7112d9	bottom-back-left	100.0	2.1	11.2
7112ec	bottom-back-right	100.0	1.4	11.8
3f0593	top-front-left	99.6	2.4	11.3
3f0604	top-front-right	100.0	2.5	10.7
3f0606	top-back-left	99.2	2.3	11.1
3f0595	top-back-right	99.2	2.4	11.5
3f0605	middle-front-left	82.1	20.9	5.6
3f065e	middle-front-right	81.3	22.3	5.5
3f0603	middle-back-left	77.5	22.9	5.1
3f0607	middle-back-right	79.8	22.7	5.2

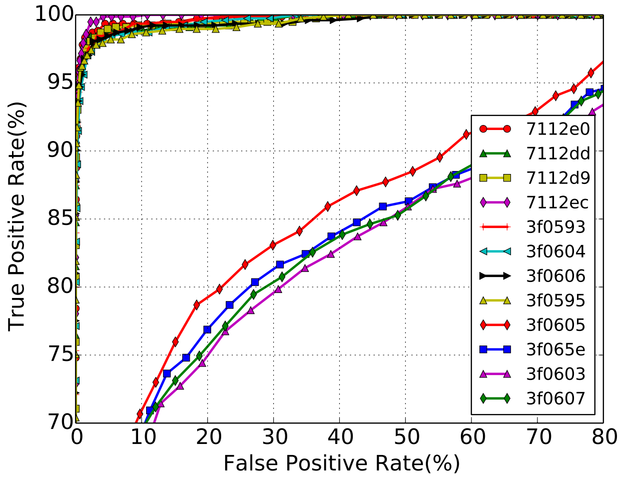
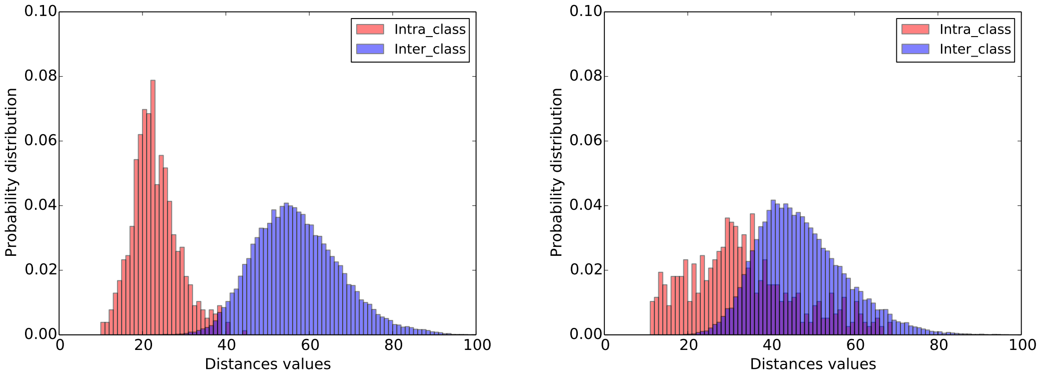


Figure 3.18: ROC curves of the GEI for different camera viewpoints.



(a) Far camera

(b) Middle camera

Figure 3.19: Intra/Inter-class variation using real silhouettes from a (a) far-bottom camera and (b) middle camera.

3.3.3 The impact of using synthetic silhouettes on the GEI performance

In this experiment, we measured the recognition performance of the GEI using synthetic silhouettes to see what effect the 3D reconstruction and projection processes have on the recognition performance. For each test viewpoint, 3D volumetric sequences were reconstructed using gait sequences from all the cameras except the test (target) camera. The volumetric sequences were then projected onto the test camera viewpoint and the resulting synthetic silhouettes were used in both the gallery and probe sets. After that, the normalized GEI was calculated as a gait feature. The recognition results are reported in Table 3.4. It can be observed that the results have similar behaviours as those obtained using the real silhouettes from the cameras. However, a slight drop in the performance is seen which might be attributed to the 3D reconstruction and projection errors. The maximum reduction in recognition rate was 0.8% from the far-ends cameras and 1.6% from the middle cameras; whilst there was no noticeable change in the values of EER and the decidability index. The ROC curves and class distributions are not displayed due to their similarities with those obtained in the previous experiment.

Table 3.4: Results of the GEI performance using synthetic silhouettes.

Camera ID	placement	CCR(%)	EER(%)	Decidability
7112e0	bottom-front-left	99.6	1.6	11.9
7112dd	bottom-front-right	99.6	1.7	11.7
7112d9	bottom-back-left	99.6	1.4	11.2
7112ec	bottom-back-right	99.6	1.2	11.9
3f0593	top-front-left	99.6	2.2	11.3
3f0604	top-front-right	99.2	2.2	10.6
3f0606	top-back-left	99.2	2.2	11.1
3f0595	top-back-right	99.6	1.4	11.8
3f0605	centre-front-left	81.0	20.7	5.6
3f065e	centre-front-right	82.9	21.5	5.7
3f0603	centre-back-left	75.9	22.9	5.3
3f0607	centre-back-right	79.1	22.1	5.1

3.3.4 Matching 3D volumetric data against single viewpoint gait data

In this experiment, we investigated the performance of matching 3D volumetric data in the gallery set with the real silhouettes from a single camera in the probe set. To do the matching, the volumetric sequences were projected into the probe image to generate synthetic silhouettes from the viewpoint of the probe. The synthetic silhouettes were computed as in section 3.3.3. We then measured the performance from each viewpoint in the Tunnel by comparing the GEIs of synthetic silhouettes with the GEIs of real silhouettes from the target camera. The recognition results are summarised in Table 3.5. It is apparent that the results are worse than those obtained using only real silhouettes or synthetic silhouettes, especially for the middle cameras. The average recognition rate was about 97% for the far-ends cameras and approximately 41.6% for the middle cameras. A significant difference was also recorded for the values of EER and the decidability index in the two sets of cameras. The ROC curves are only shown for the four middle cameras in Fig. 3.20. These curves reflect the poor recognition performance from these cameras. The class distributions are plotted in Fig. 3.21 for one camera in each set. In this figure, the overlap between intra- and inter-class distributions for the middle camera reaches its highest level. Furthermore, the intra-class distribution for the middle camera is wider than that obtained from the far-end camera. This means that the ability to identify people using the GEI from the middle wide-angle cameras is lower than that using the GEI from the far-ends cameras.

A similar analysis was done by Seely in his PhD thesis [98], showing that recognition was impossible using the four middle cameras due to the difficulty in obtaining accurate camera calibration information because of their significant lens distortion. Therefore, these cameras were excluded from the analysis and the results were only reported for the four far-top cameras. An average recognition rate of about 79.5% was recorded for the front-view cameras and nearly

53.8% for the rear-view cameras [98]. Our results reported in this section are better than those achieved by Seely. The better quality of our camera calibration is the reason for the difference in performance as the quality of calibration affects the appearance of the synthetic silhouette generated from the 3D volume.

Table 3.5: Recognition performance for matching 3D volumetric data against single viewpoint data.

Camera ID	placement	CCR(%)	EER(%)	Decidability
7112e0	bottom-front-left	96.1	4.4	9.5
7112dd	bottom-front-right	97.2	4.9	9.0
7112d9	bottom-back-left	97.2	5.6	8.5
7112ec	bottom-back-right	97.6	3.1	10.1
3f0593	top-front-left	96.5	4.1	9.1
3f0604	top-front-right	95.7	4.6	8.7
3f0606	top-back-left	96.8	3.4	9.4
3f0595	top-back-right	98.4	3.1	10.3
3f0605	centre-front-left	48.1	27.1	3.5
3f065e	centre-front-right	53.8	25.9	3.4
3f0603	centre-back-left	30.2	33.4	2.4
3f0607	centre-back-right	34.5	29.9	2.8

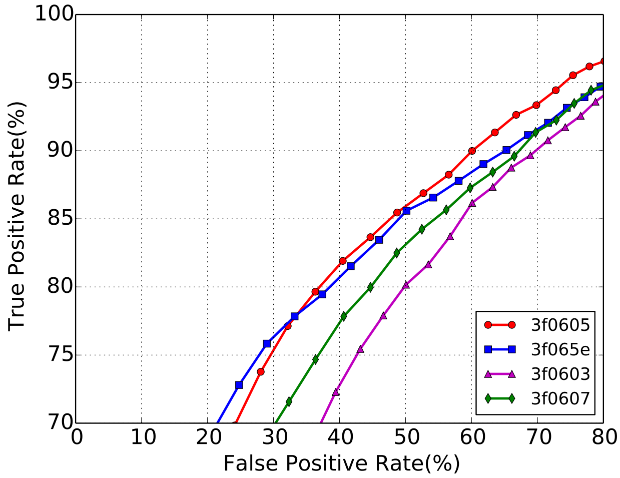
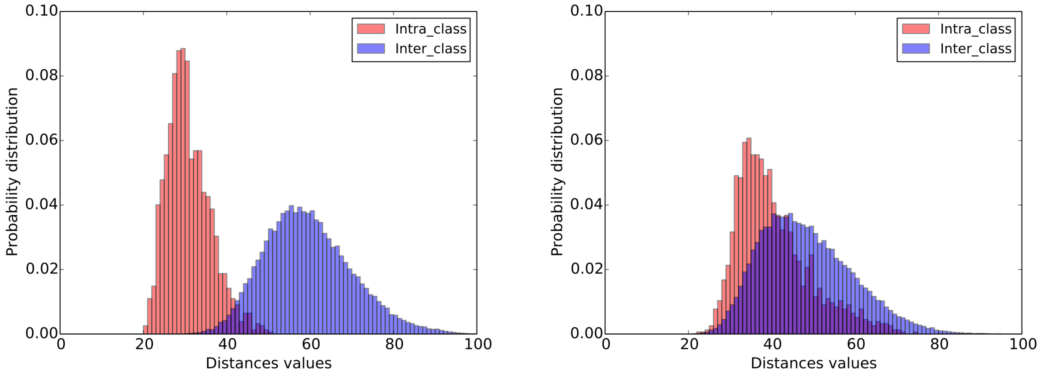


Figure 3.20: ROC curve for matching 3D volumetric data against 2D silhouettes.

3.4 Discussion

In this chapter, the basic stages of matching 3D volumetric data against 2D silhouettes from an independent (arbitrary) view are explained. Several factors that affect the matching performance have been analysed through a number of experiments. The first experiment showed that using different camera configurations did not influence the performance as the same type of data are used in both the gallery and probe. However, omitting the middle cameras in the Tunnel



(a) Far camera

(b) Middle camera

Figure 3.21: Intra/Inter-class distributions for matching 3D volumetric data against gait data from an arbitrary camera viewpoint: (a) far-bottom camera and (b) middle camera.

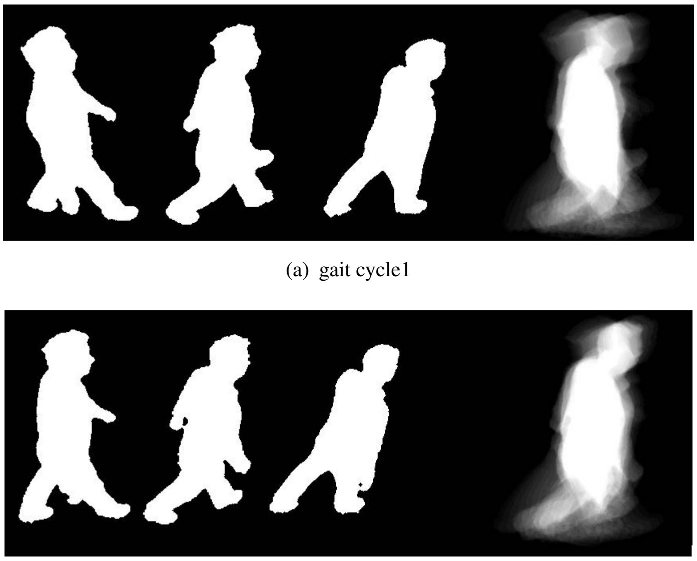
significantly distorted the quality of the 3D reconstructed volumes while no noticeable distortion happened when other cameras in the Tunnel were omitted. This emphasises the importance of including the middle cameras in order to obtain better quality 3D volumetric data.

In the second experiment, we analysed the discriminatory power of the GEI using real silhouettes from different camera settings. The results showed a high discriminatory power of the GEI for silhouettes from the far cameras with a small field-of-view (weak perspective silhouettes) and a low discriminatory power of the GEI from the middle (near) cameras with a wide field-of-view (strong perspective silhouettes). The visual inspection of the silhouettes from the far cameras showed that the appearance, orientation and observation angle for the silhouettes in one gait cycle did not change significantly and thus the temporal variation of the silhouette shape can efficiently be captured by the GEI. On the other hand, the silhouettes from the middle perspective cameras exhibit a variation in the appearance, orientation and observation angle within one gait cycle as shown in Fig. 3.22. In this figure, several silhouettes are shown from two different gait cycles for the same subject with their GEIs. The GEI cannot efficiently capture rapidly changing shapes from a perspective camera as the relevant body parts cannot be matched accurately. Also, these two gait cycles start at two different positions in the Tunnel, which result in two different GEIs. To further illustrate that, we drew the true positives [2] and false negatives [3] for this experiment using the position of the centroid along the Z-axis for the middle frames in all 3D gait cycles in Fig. 3.23. As can be seen that the positively (correctly) matched GEIs occur when the compared gallery and probe gait cycles are in a nearly same position in the Tunnel; while the GEIs are negatively matched (i.e. failed to match) when the positions of the compared gallery and probe gait cycles are different.

The results presented in the third experiment show that there is a slight drop in the recognition performance using synthetic silhouettes. This drop might be attributed to the 3D reconstruction

[2] True positives are the GEIs of the same subjects which correctly match.

[3] False negatives are the GEIs of the same subjects which fail to match.



(a) gait cycle1

(b) gait cycle2

Figure 3.22: Several frames in two gait cycles for the same subject from one of the middle cameras in the Tunnel.

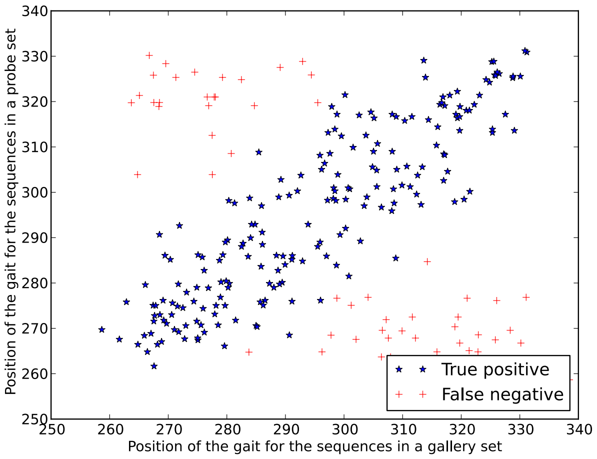


Figure 3.23: The positions of gait cycles along the Z-axis (in cm) for true positives and false negatives of matching samples from a middle camera.

and projection errors. The last experiment reveals that matching 3D volumetric data against 2D silhouettes from an arbitrary view produced a high recognition performance from the far cameras and a poor performance from the middle perspective cameras. There are two possible reasons for the poor performance by the middle cameras. The first reason is a distortion that might happen in the 3D volumetric data when omitting one of the middle cameras in the Tunnel as demonstrated in section 3.3.1, which could result in a discrepancy between the shape of the synthetic silhouettes from 3D volumes and real silhouettes. The second reason is a lower discriminatory power of the GEI for the silhouettes from the perspective cameras as explained in section 3.3.2.

These results are still promising and illustrate that matching 3D volumetric data against an arbitrary view(s) is possible. Based on these results, it was decided to do the following work to improve the matching performance.

1. All twelve cameras in the Tunnel will be used for a better quality of the 3D volumetric data for our gallery.
2. To do recognition from an arbitrary view(s), an independent camera(s) will be added to the Tunnel to capture 2D silhouettes of the subjects for our probe.
3. An improved 3D to 2D matching algorithm should be developed to ensure that the silhouettes in each pair of the gallery and probe gait cycles share the same appearance, orientation and observation angles.
4. An efficient feature extraction and matching procedure should be used to deal with the variations in the orientation and appearance of silhouettes in one gait cycle.

This work is described in the following chapters.

Chapter 4

Tunnel Setup and 3D-2D Gait Dataset Acquisition

The findings in chapter 3 showed that matching 2D silhouettes from a single camera(s) in the Tunnel with the 3D volumetric data can be efficiently performed for the 2D silhouettes from the narrow field-of-view cameras which are placed far from the walking subject. These findings also highlighted the issues that affect the recognition using cameras near to the subject. The first issue is the omission of one of the middle cameras to build the probe set from an independent view. This might adversely influence the accuracy of the 3D volumetric data and create a discrepancy between the synthetic silhouettes generated from the distorted 3D volumes and the real silhouettes from the camera. The second issue is the matching of the 3D with 2D gait data regardless of the distance between the subject and the camera. The distance from the camera has a significant effect on the appearance of the subject in a 2D image. A new algorithm for matching 3D with 2D gait data was developed to overcome this issue. The details are given in the next chapter.

To tackle the first issue, all the cameras in the Tunnel are used to improve the quality of the 3D reconstructed data to build the gallery set. Two asynchronous perspective cameras are also added to capture silhouette images from two independent (arbitrary) views to build the probe set. Initial tests for the multi-view capturing system in the Tunnel revealed some problems as the original system built by Seely [98] malfunctioned and was replaced with a modified version without any testing. As explained before, the original Tunnel system consisted of four computers to store gait images from the twelve synchronous cameras. In addition, one efficient computer was used to compute the 3D reconstruction and manage the capturing process from the whole system. However, the modified Tunnel system includes only one efficient computer where all the twelve cameras were directly connected to it through four hubs. The LED lights and the two break-beam sensors, which were used to activate the start and stop of the capturing process have been disconnected. The problems in the Tunnel had been investigated and proper solutions had been devised to ensure the reliability and consistency of the system. This was not a trivial job as several alterations to the modified system's software and hardware were made. A poor illumination problem in the Tunnel was investigated and additional lighting sources were added. The alignment of the twelve synchronous cameras was explored and new positions and orientations of the cameras were devised. The work in the Tunnel took approximately three years, which delayed the collection of a large gait dataset. The detailed description of the main modifications and expansions to the layout of the Tunnel will be given in this chapter.

4.1 Extensions to the Biometrics Tunnel configuration

The aim of the work described in this section is to provide statistically view-independence for the probe set. Two wide-angle lens cameras were added to the Tunnel to provide gait images from two arbitrary viewpoints. The positions and orientations of these cameras were chosen such that the first (Point Grey Flea) camera was placed near one end of the Tunnel to capture a subject's gait from the rear (a transition of views from an approximately oblique to a rear view) while the second (Dragonfly) camera was mounted on one of the walls near the middle of the Tunnel to capture a subject's gait from the side (a transition from an approximately side to an oblique view). Fig. 4.1 shows the placement of the two cameras in the Tunnel. The first camera (rear) had a resolution of 1024×768 pixels and was placed at a height of 175cm from the floor; pointing inwards. The second camera (side) had a resolution of 640×480 pixels and was mounted at a height of 236cm; pointing downwards. Both cameras had a focal length of 4mm and captured video footage at a rate of thirty frames per second. Initially, only the rear camera was connected to its dedicated computer via IEEE1394 bus. FlyCapture software was installed on this computer to enable the acquisition of raw video footage from the camera. Using this software, the user can specify the camera settings (e.g. shutter speed and gain) for each session and control the starting and stopping of the capturing process. The side camera was not connected at the beginning until the validity of gait images captured by the rear camera had been verified and the status of the entire system had been checked.

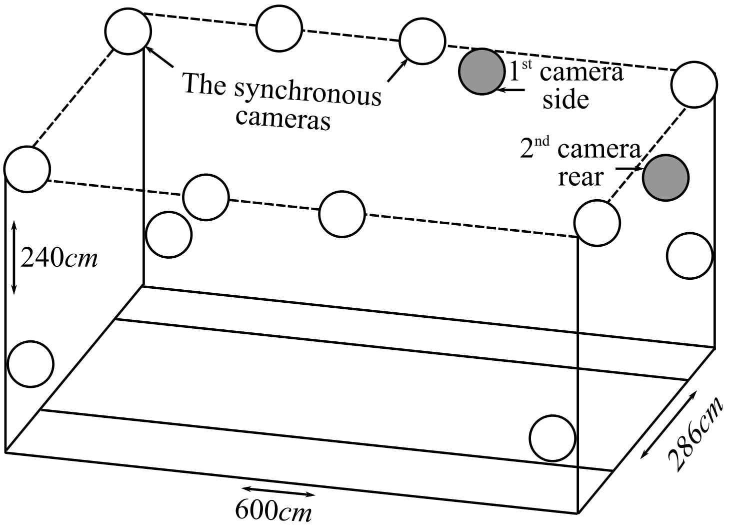
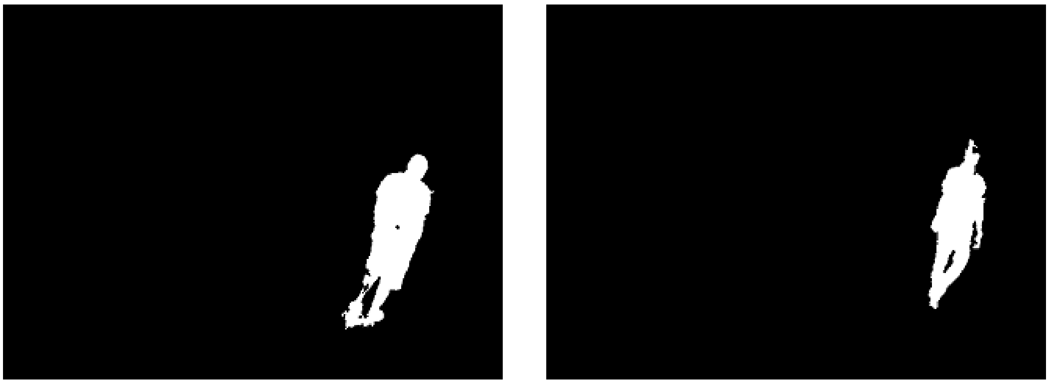


Figure 4.1: The placement of the two arbitrary cameras in the Tunnel.

Once the Tunnel system had been expanded, an initial test was carried out by collecting few samples from the far camera to verify the reliability of the acquired data. Ten subjects were recorded: 4 female and 6 male. Each walked six times along the central line of the Tunnel from one end to the other in one direction. This resulted in a total of 60 single-view gait sequences. As no break-beam sensors were connected to the system at that stage, several empty and redundant frames were included in each walking sequence. To reduce the storage requirements and the time required to process each foreground sequence, the empty frames were detected and removed. The following procedure was implemented: two white rectangular masks were determined to mark the start and end of the active walking path in the Tunnel and two corresponding threshold values were set. Then, a simple background segmentation algorithm was implemented by choosing the first empty frame as a background model. Each subsequent frame was later subtracted from the background and compared against a predetermined threshold to separate a walking subject. After that, the first and second masks were used to isolate regions in the segmented image that represent the start and end of the walking path. The area within each region was measured and compared against the first and second thresholds to mark the first and last valid foreground frames. Finally, we visually inspected the extracted frames to check the case of incomplete walking sequences.

Next, the silhouettes were extracted from each sequence using the background segmentation algorithm described in section 3.2.1.1. We noticed severe background segmentation errors in the extracted silhouettes. The silhouettes were retouched manually to improve the quality of segmentation. However, shadow effects around the feet of the subject, and some missing areas in the head and leg regions still existed as shown in Fig. 4.2. A possible cause for those was poor illumination especially at both ends of the Tunnel. This issue is further illustrated in the next section. Another issue was that several frames were missing in each sequence, which made it difficult to detect a complete gait cycle. The omission of the frames was due to improper camera settings and an insufficient storage space available on the computer. The proper settings for the camera were then determined and one 2TB hard-disk was added to the computer to enable the collection of a substantial amount of data.



(a) Shadow around the feet region

(b) Missing area in the head region

Figure 4.2: Examples of poor segmentation results.

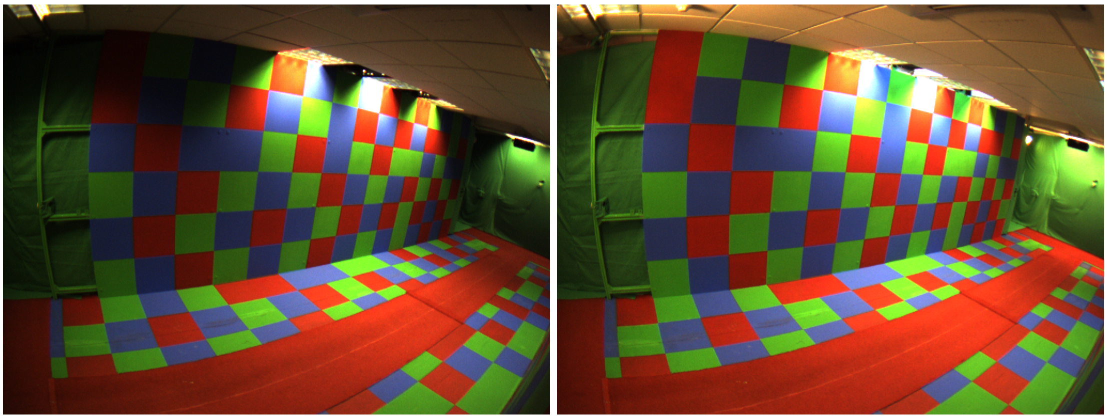
4.2 Improvement of the Biometric Tunnel Condition

The visual inspection of the segmented silhouettes from the collected samples obtained as described in section 4.1 revealed some background segmentation errors which could be caused by poor illumination and improper settings of the camera. The light sources were insufficient to illuminate all the areas in the Tunnel and were not distributed properly throughout the Tunnel. The illumination at the start and end of the walking path was poor in comparison with the middle of the path. This resulted in bright and dark areas in the Tunnel, which create noise and shadow effects in the extracted silhouettes as illustrated in Fig. 4.2. Several attempts were made to improve the illumination inside the Tunnel using a variety of light sources in different areas.

One proposed solution was to distribute light sources in the middle of the ceiling along the length of the walking path to evenly illuminate the whole region and provide comfortable lighting for walking subjects. However, this solution was impossible due to the construction of the ceiling tiles. A second proposed solution was to place four LCD lights at the far corners of the Tunnel. However, it was found that those lights were unsuitable due to their flickering effect. A suitable solution to provide a compromise between the quality of illumination and the comfort of the participants was to use four point lights at the four top corners and two studio light sources at the entry and exit of the Tunnel. Fig. 4.3 shows the Tunnel before and after additional lighting. Having a sufficient illumination in the Tunnel made it easier to choose suitable values for shutter speed and gain for the cameras to reduce the effect of motion blur in captured images and improve the signal to noise ratio.

There were some areas at both ends of the Tunnel covered by grey cloth and were clearly seen by some cameras. It was also noticed that the quality of background segmentation deteriorated as a subject passed through grey areas in the Tunnel. The grey colour was often worn by many people, which made it difficult to separate a subject from his/her surroundings. It was therefore decided to cover the grey areas with the green cloth as the green is one of the primary colours in the Tunnel and it is unlikely to be worn by many people.

With the new modifications to the Tunnel, further samples were captured from the rear camera and from the synchronous cameras to investigate further problems. No automatic way was available at that stage to start and stop the capturing process from the rear camera and the synchronous cameras. Therefore, this process was carried out manually using the corresponding computers. Five people were filmed, each walking five times in the Tunnel. A total of 25 multi-view and single-view gait sequences were collected. The silhouettes were firstly extracted from the rear camera and from the synchronous cameras. No severe background segmentation errors were noticed in the extracted silhouettes from all cameras in the Tunnel. We then built the 3D sequences of the participants using multi-view silhouettes and calibration information from the twelve synchronous cameras. The visual inspection of the 3D reconstructed data revealed significant missing parts in the 3D volumes especially in the leg region as demonstrated in Fig. 4.4. Further investigation was required to trace back the source of this problem.



(a) Before additional illumination (b) After additional illumination

Figure 4.3: Tunnel illumination condition.

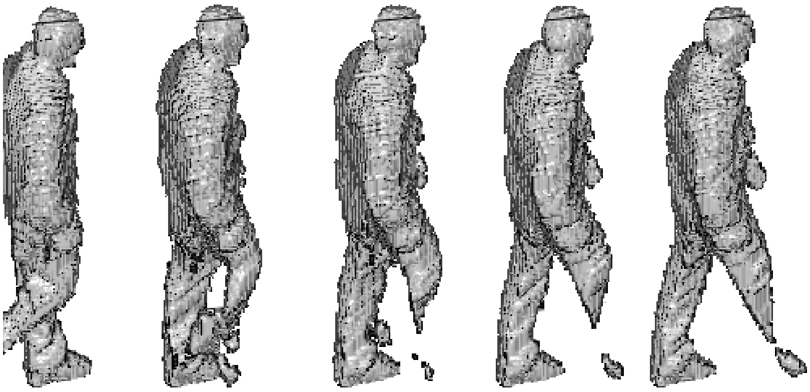


Figure 4.4: Examples of 3D reconstructed volumes with significant missing body parts.

4.3 Modifications of the multi-camera capturing system

The Biometrics Tunnel was well equipped with a multi-camera capturing software and hardware, although some issues had been revealed during the experimental work as described in sections 4.1 and 4.2. The collection of a large gait dataset was postponed until all the issues had been resolved, and the necessary reliability testing had been made. One of the issues was the missing parts in the 3D volumes captured by this system as illustrated in section 4.2. An extensive investigation was carried out to identify the source of this issue. First of all, the extracted silhouettes of the badly reconstructed volumes were visually inspected and no missing parts were noticed. Secondly, the capturing software was reviewed and it was found that the improper time synchronisation in that software allowed a delay of 60 milliseconds between the cameras and then dropped the latest frames, which had a negative impact on the quality of the 3D reconstructed volumes. In order to tackle this problem, the time synchronisation code was rectified to allow storing the latest frames and inserting the time-stamp and camera-id in the first 8 bytes of each frame. This allowed the re-synchronisation of simultaneous frames according to their time-stamps after the capturing process was completed. A tracking procedure was also implemented to monitor the status of the entire system and to detect cases of dropping frames from any camera in the Tunnel. After these modifications were made, some samples were captured to check the missing frames issue. It was found that few frames were still sometimes missing from two cameras. Several efforts were made to tackle this problem but no solution was found. It was initially believed that this issue was caused by having a large number of cameras (i.e. 12 cameras) connected to the efficient computer. One attempt was made by disconnect some of the cameras from the computer but this did not tackle the issue because the system still dropped few frames from two of the cameras. This issue was likely to be an electronics problem. To mitigate the effect of dropping frames, some efforts were spent to ensure that the dropped frames were from the far-ends synchronous cameras which have only a slight influence on the 3D reconstruction quality as illustrated in section 3.3.1. Fig. 4.5 shows an examples of several 3D volumes using the modified Tunnel system.

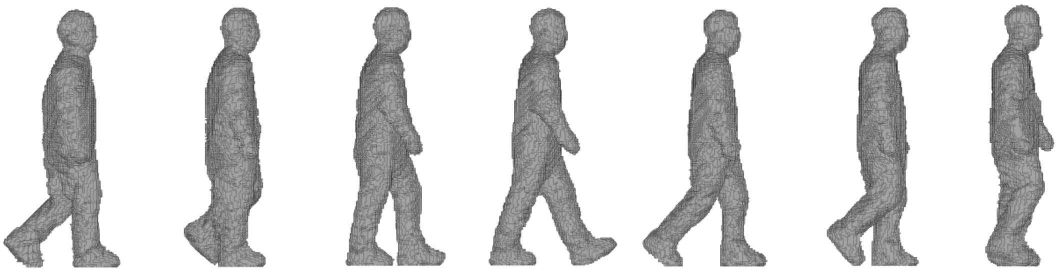
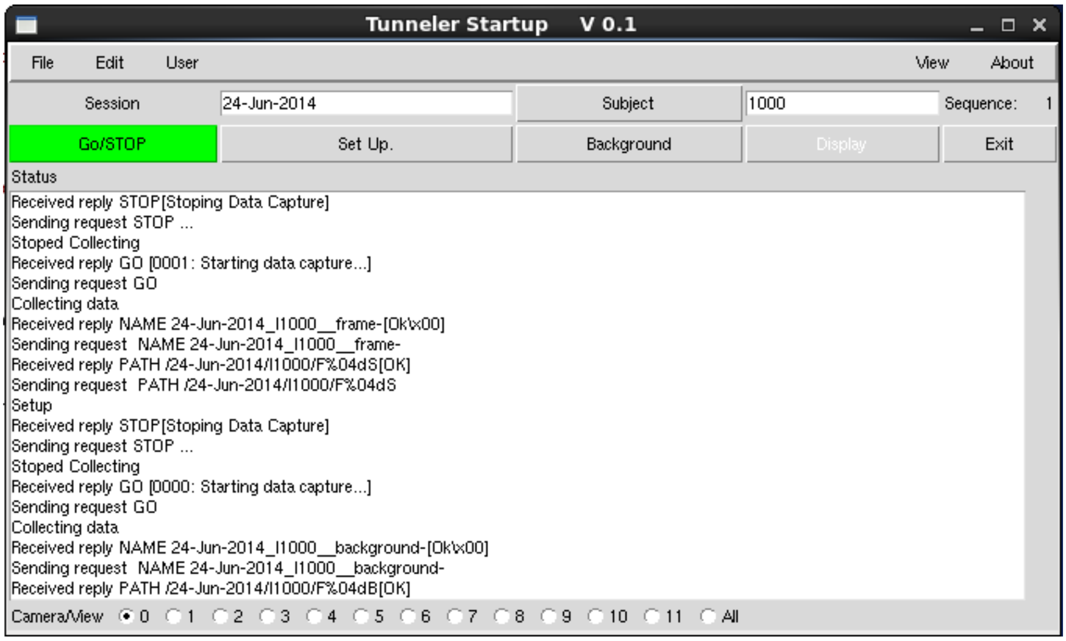


Figure 4.5: An example of 3D volumetric data using the modified Tunnel system.

The settings of the synchronous cameras were also an issue. Each time the system was restarted, these settings returned to their manufacturing values. The proper settings for each camera had to be determined manually each time the system was restarted. This task was time-consuming and had unreliable because of the number of cameras in the Tunnel. In the revised version of the capturing software, the proper settings were determined according to the quality of the segmented silhouettes. Then, the new settings were automatically set up for all the cameras at the beginning of each session. To start and stop the key operations and manage the capturing process from the synchronous cameras, a Graphical User Interface (GUI) was written using the Python programming language to allow the supervisor to start a new session, assign subject-id and specify the background and foreground modes. It also allowed the user to display a live image from any synchronous camera in the Tunnel to interactively adjust the camera focus and pose. Fig. 4.6 shows a screen-shot of the GUI when capturing the first sequence for subject-id (1000).

Another issue that arose during the experimental work was the alignment of the twelve synchronous cameras. It was noticed that the old placement and orientation of these cameras did not cover a sufficient common area in the Tunnel. Several attempts were made to find the best alignment for the twelve cameras. Unfortunately, it was difficult to align the four bottom cameras



Session 24-Jun-2014 Subject 1000 Sequence: 1

Status

```
Received reply STOP[Stoping Data Capture]
Sending request STOP ...
Stoped Collecting
Received reply GO [0001: Starting data capture...]
Sending request GO
Collecting data
Received reply NAME 24-Jun-2014_I1000__frame-[Ok\x00]
Sending request  NAME 24-Jun-2014_I1000__frame-
Received reply PATH /24-Jun-2014/I1000/F%04dS[OK]
Sending request  PATH /24-Jun-2014/I1000/F%04dS
Setup
Received reply STOP[Stoping Data Capture]
Sending request STOP ...
Stoped Collecting
Received reply GO [0000: Starting data capture...]
Sending request GO
Collecting data
Received reply NAME 24-Jun-2014_I1000__background-[Ok\x00]
Sending request  NAME 24-Jun-2014_I1000__background-
Received reply PATH /24-Jun-2014/I1000/F%04dB[OK]
```

Camera/View ● 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10 ○ 11 ○ All

Figure 4.6: A screen-shot of the GUI that was used to manage the acquisition of gait images from the synchronous cameras.

with the 8 top cameras. New wide-angle lenses were purchased for the bottom cameras to allow a sufficient coverage for the area in the Tunnel and allow easy alignment. The mounting brackets holding the cameras were also tightened to minimise the possibility of a camera movement. Then, the cameras were fully re-calibrated. The old hubs that connected the cameras to the efficient computer were also replaced with the new ones after they failed.

After that, the additional side camera was connected to its dedicated computer. A FlyCapture software was installed on this computer to enable the acquisition of raw video footage from this camera. The settings of the camera were carefully chosen to reduce the effect of motion blur and background segmentation errors. Several samples were captured from this camera to check their validity. Finally, the two independent cameras were calibrated and their external (position and orientation) and internal parameters were calculated using the same procedure described in section 3.2.1.2. The modified Tunnel system at that stage included one powerful computer to manage the acquisition of gait images from the synchronous cameras and two dedicated computers to store gait images from the two independent cameras.

4.4 A 3D-2D Gait dataset collection procedure

After the proper settings for all cameras in the Tunnel were determined and the necessary reliability testing had been finished, a small and a large gait datasets was collected to be used in the analyses. The small gait dataset was used to analyse the performance of the matching algorithm in chapter 5, while the large gait dataset was used for performance analyses of the methods in chapters 6 and 7. During the capture process, the background samples were recorded for each subject before capturing his/her walking sequences to reduce the effect of variations in the lighting between background estimation and segmentation. The people recorded in these datasets wore their normal clothes and walked in their normal speed to simulate real gait data. Participation in the gait experiments was voluntary, and no financial incentive was introduced. Upon the arrival of the participant, the supervisor demonstrated the experiment's aims, aspects and the procedure to be carried out by the participant. To explain the walking procedure, the supervisor conducted a trial walk through the Tunnel. After that, the capturing process started according to the supervisor instructions.

During collecting the smaller gait dataset, single-view gait sequences were captured from the two independent cameras using the FlyCapture software in the two dedicated computers connected to these cameras. Multi-view gait sequences were recorded using the multi-camera capturing software installed in the powerful computer and this operation was controlled via the GUI. The supervisor instructed to each participant when to start and stop walking in the Tunnel. The capturing process was started and stopped manually on each of the three computers in the system.

It was found from collecting the small gait dataset that the time required to record a subject's gait images from the independent and synchronous cameras was long (about 30 minutes) since the management of the capturing process was distributed on three computers, which was troublesome, time-consuming and subjected to human errors. Therefore, several modifications were made to provide a central management for the capturing process from all the cameras. These modifications facilitated the collection of the larger gait dataset. The two computers connected to the independent cameras were replaced with a single efficient computer. The multi-camera capturing software for collecting gait images from the synchronous cameras was modified and used to acquire gait images from the two independent cameras. The break-beam sensors, which were located at the entrance and exit of the Tunnel, were re-connected to the current system. Two red LED lights were also placed at both ends of the Tunnel to guide the participants during the experiment. During this period, the pose of some cameras including the two independent cameras was changed due to some technical work inside the Tunnel. The cameras were re-localised and re-calibrated before collecting the second dataset.

A Raspberry Pi computer was added in which a new version of the GUI was written to enable the supervisor to manage the capturing process on all computers. This version had three modes of operation. In the first mode, recording a walking sample(s) was done through start/stop bottoms The second mode allowed the supervisor to record a number of frames over a specific time period. In the third mode, a number of walking samples for a subject were specified and captured such that starting and stopping different walks were controlled by trigging the break-beam sensors at both ends of the Tunnel. Fig. 4.7 shows the connection of the Raspberry Pi computer in the Tunnel. Several trial walks were recorded and checked for their validity before collecting the larger gait dataset. For this dataset, the second mode was selected to record background frames whilst the third mode was used to acquire walking samples for each subject. When the red light turned off, the participant walked several times through the Tunnel by triggering the break-beam sensors.

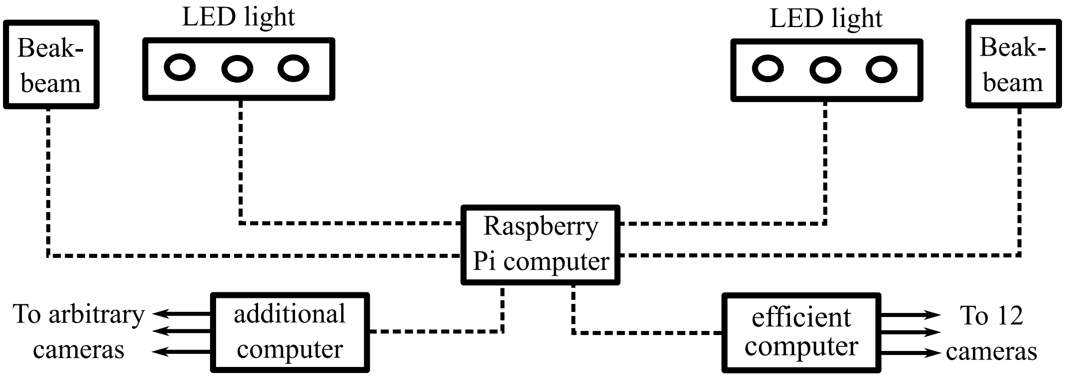


Figure 4.7: Connection of a Raspberry Pi computer in the Biometrics Tunnel.

After completing the capturing process, the gait images from the synchronous cameras were re-synchronised using their time-stamps to build 3D volumetric sequences. All the processing algorithms were written in the Python programming language because of its flexibility. The details of the two collected gait datasets are given in the next two chapters.

4.5 Discussion

Several problems were discovered in the modified Tunnel system. These problems were investigated and proper solutions were devised. The main problem was the inaccurate time synchronisation, which allowed a delay of 60 milliseconds between the cameras and also dropped the latest frames. This problem caused missing parts in the 3D reconstructed volumes of people. A modified version of the multi-capture software was written to rectify this problem by storing the latest frames and inserting the time-stamps at the beginning of each frame. This allowed to re-synchronisation of the frames after completing the capture process. Another problem was the alignment of the synchronous cameras. The alignment determines the common area in the Tunnel that can be seen by the cameras. It could also affect the way the back-projected rays from the cameras intersected to build the 3D shape of a subject. Therefore, new poses were devised for all cameras to achieve better alignment. Furthermore, a poor illumination condition was investigated where there were dark and bright areas in the Tunnel. A possible solution for the illumination problem was to include additional lighting at the ends of the Tunnel to reduce shadow and noise effects. This improved the background segmentation. The layout of the Tunnel was also expanded by adding two asynchronous cameras to capture 2D silhouettes from two independent viewpoints. The new version of the Tunnel was able to capture 3D volumetric sequences as well as independent 2D gait sequences without the need to repeat the capture process for each type.

Chapter 5

Gait Recognition with Large Perspective Distortion

A poor recognition performance was reported in chapter 3 using silhouettes from the wide-angle lens cameras in the middle of the Tunnel where a subject is close to the camera. In this case, we noticed that there were a large changes in the appearances of the silhouettes in one gait cycle. Using a global statistic (e.g. GEI) for feature extraction and direct matching of gait data in the gallery and probe sets affect the performance of gait recognition. Motivated by these, we proposed a new gait recognition technique to identify subjects captured by an arbitrary perspective camera mounted on the wall in a constrained narrow corridor using 3D models (volumes) of people. The main contribution in this technique is the elimination of the effect of appearance variations due to position changes when considering gait data captured by cameras with different settings. To simulate walking in a narrow corridor, a small gait dataset was collected for which two new independent cameras with wide-angle lenses were placed at two different positions in the Tunnel, as explained in chapter 4, to record people walking on a straight line from the entrance to the exit. Extensive analyses were conducted to show the effectiveness of matching 3D volumes with perspective distorted silhouettes from independent (arbitrary) views.

5.1 The main matching technique

The first stage is to enrol people using a set of synchronous cameras to reconstruct their 3D volumes for the gallery set and capture gait images of walking subjects from an independent camera to build the probe set. Then, for each pair of 3D and 2D gait sequences in the gallery and probe, one gait cycle is detected and a leading foot in each of these cycles is determined. The proposed technique mainly depends on a 3D alignment and projection process to allow synthetic silhouettes from 3D volumes and real silhouettes from the camera in one gait cycle to share the same appearances, orientations and observation angles. This is done by estimating the position of each pair of 3D and back-projected 2D gait cycles in 3D space to determine the amount of displacement between them. All 3D volumes in one gait cycle are then aligned and projected onto the 2D image of the probe camera. After that, gait features are extracted from the synthetic silhouettes and real silhouettes using Generic Fourier Descriptors and compared using Dynamic Time Warping. Finally, a subject is recognised based on a minimum distance between gait features. Fig. 5.1 shows the key operations in the proposed technique. The details of these operations will be given in the following subsections.

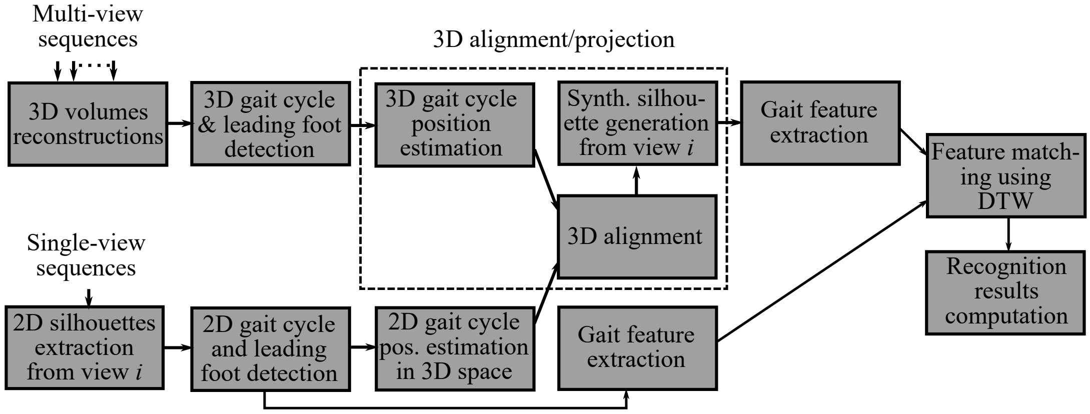

Multi-view sequences
3D alignment/projection
3D volumes reconstructions
3D gait cycle & leading foot detection
3D gait cycle position estimation
Synth. silhouette generation from view *i*
Gait feature extraction
Feature matching using DTW
3D alignment
Recognition results computation
Single-view sequences
2D silhouettes extraction from view *i*
2D gait cycle and leading foot detection
2D gait cycle pos. estimation in 3D space
Gait feature extraction

Figure 5.1: The stages of the proposed technique.

5.1.1 Gait cycle and leading foot detection

The proposed technique starts by detecting one gait cycle from each 3D and 2D gait sequences in order to compute the matching. It is also essential to detect the leading foot of a person in each gait cycle so that the starting and ending phases of each pair of 3D and 2D gait cycles are aligned. The gait cycle and leading foot detection will first be explained for a 3D gait sequence and then for a 2D sequence.

• Gait cycle and leading foot detection for a 3D gait sequence

The gait cycle is determined from each 3D gait sequence using the variation of the 3D bounding box width along the direction of walking as explained in section 3.2.2. One 3D gait cycle is labelled using three consecutive volumes when the two legs completely overlap (mid-stance). Then, the leading foot in a 3D gait cycle is detected by placing a 3D bounding box around the active region in the first volume where the two feet are farthest away from each other (double-support stance) and dividing it horizontally into two sections: upper and lower. The upper section occupies 80% of the volume height while the lower section, which houses the two feet, represents only 20%. Then, a vertical plane is used to equally sub-divide the lower section into two segments as illustrated in Fig. 5.2. We noticed sometimes that the two feet cannot be separated using a single plane. Therefore, we shifted the vertical plane by 5 voxels towards the positive x-axis to determine the left foot region and by 5 voxels towards the negative x-axis to define the right foot region. These numbers have been determined by a visual inspection of several 3D gait cycles. After that, the position of the most extreme voxel with the highest value along the z-axis in these regions is labelled the leading foot in the 3D gait cycle. In this thesis, only one gait cycle was determined for each 3D volumetric sequence, which starts with either the right or the left leading foot (i.e. no constraint is imposed on the selection of the leading foot in 3D gait cycles).

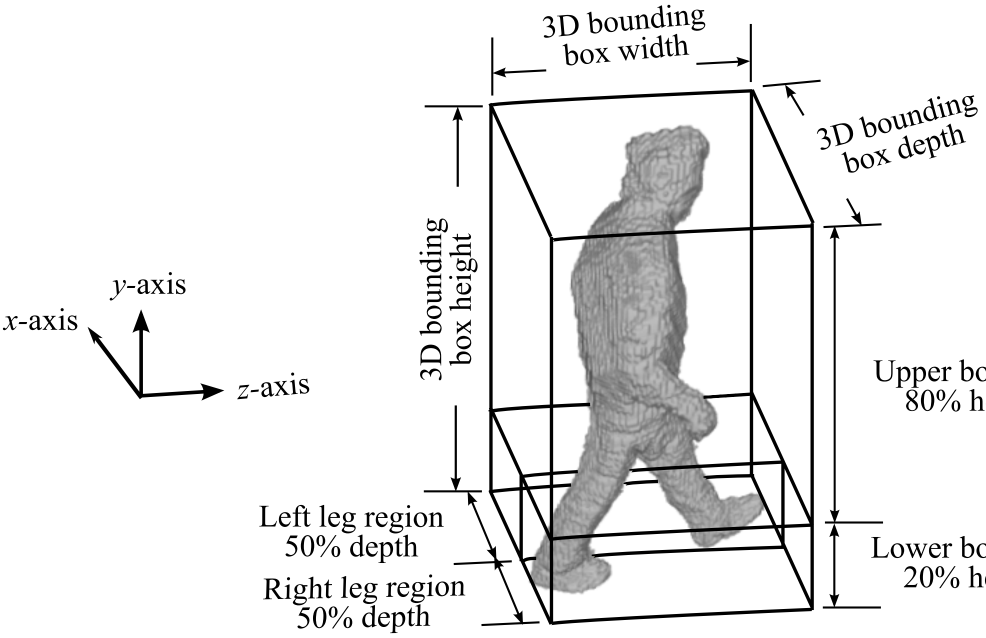

3D bounding box width
3D bounding box depth
3D bounding box height
y-axis
x-axis
z-axis
Upper body part 80% height
Lower body part 20% height
Left leg region 50% depth
Right leg region 50% depth

Figure 5.2: Detecting a leading foot in a 3D volume.

• Gait cycle and leading foot detection in a 2D gait sequence

Detecting a gait cycle in a 2D gait sequence is not a trivial task because of the way the camera observes the movement of the feet from different viewpoints and the effect of perspective distortion on the captured images. The variation of the bounding box width can usually be used to determine the gait cycle from an approximately side-view sequence since the extension and contraction of the projection of the two feet can clearly be seen from this viewpoint. The first Affine Moment Invariant (AMI) [28] was also calculated in [45] to label gait cycles in side-view gait sequences. To derive the first AMI, the central moments [28], which are region-based shape descriptors, are first calculated from the image. The formula for the central moment of order (p, q) is

$$\mu_{pq} = \sum_{u=1}^{M} \sum_{v=1}^{N} (u - u_c)^p (v - v_c)^q F(u, v) \tag{5.1}$$

(u_c, v_c) is the centroid of the shape, $F(u, v)$ is the image intensity at pixel (u, v) and $M \times N$ is the image dimension. The first AMI [28] is then calculated from

$$I_1 = \frac{1}{\mu_{00}^4} (\mu_{20}\mu_{02} - \mu_{11}^2) \tag{5.2}$$

On the other hand, the bounding box height is used for gait cycle detection in front-view sequences [17]. Several studies [5, 57, 58] used the aspect ratio (i.e. width/height) signal of a moving person over time for gait cycle estimation from different viewpoints. For perspective silhouettes collected in the Tunnel which include a change in observation angles in each gait sequence, we found that the first AMI gives more stable results for gait cycle estimation as shown in Fig. 5.3. In this figure, the variation of the bounding box width, height, aspect ratio and the first AMI over time are computed and smoothed using Gaussian filter of a sigma 1.5, to reduce the effect of outliers, for a gait sequence captured by the rear camera in the Tunnel. The curve produced by the first AMI contains several peaks and valleys that relate to the double support and mid-stance phases during walking. The detection of three consecutive valleys is labelled one gait cycle. The performance of the bounding box approaches is subject to errors which could be due to the orientation and observation angle variation in the same sequence, while the invariant properties of the AMI could better cope with these variations.

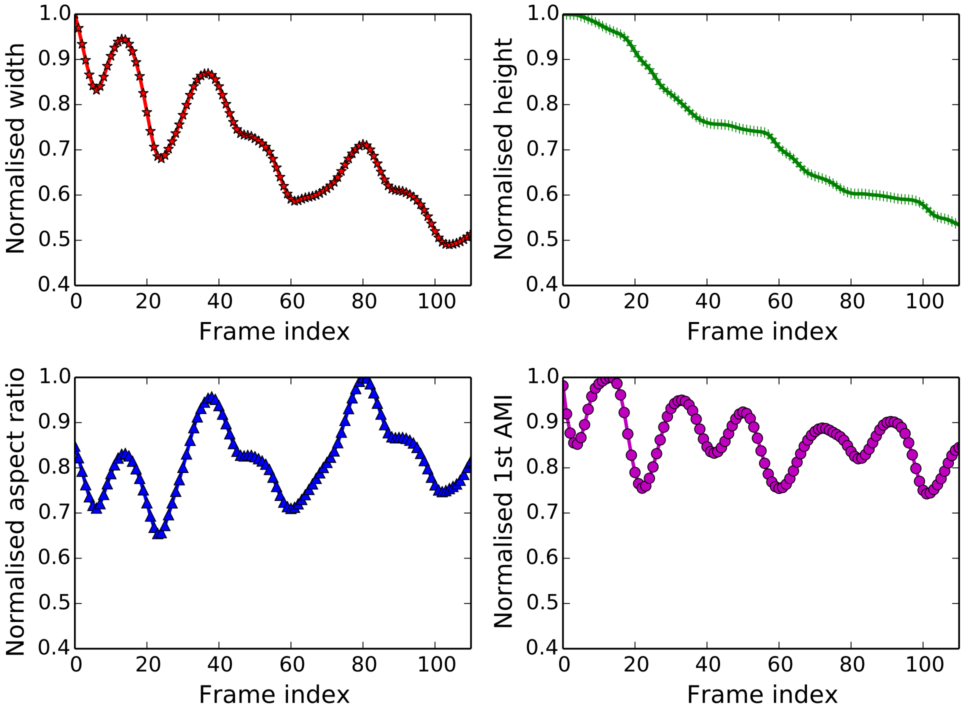


Figure 5.3: 2D Gait cycle estimation for a silhouette sequence with a perspective distortion using different approaches (bounding box width, height, aspect ratio and the first AMI).

After one gait cycle has been determined from each 2D gait sequence, the leading foot in the gait cycle should be detected. This step also depends on the nature of the gait data and the viewpoint of the camera. The leading foot can be detected from the first frame at double support phase of a gait cycle. For gait cycles captured from a nearly side-view, the perspective projection property of a camera would make the closer foot to the camera lower than the far foot. Based on the direction of walking with respect to the camera (i.e. from a left to right or right to left), the leading foot could be detected. In a case of a front-view gait cycle, the leading foot would be lower than the latest foot. For perspective silhouettes captured by the two independent cameras in the Tunnel, we used the following procedure according to the motion of the two feet and the nature of the silhouettes: a bounding box is placed around the body area in the first frame where the two feet are farthest away (double support), which is then divided into two vertical sections of predefined sizes as shown in Fig. 5.4. The upper section occupies 80% of the bounding box height and the lower section accounts for only 20%. This division is chosen based on a visual inspection. From the selected viewpoints in the Tunnel, when the left foot is in front it would be partially occluded by the motion of the right foot (the closer foot to the camera) and the lower section of the bounding box would include only one single connected component that refers to the right foot. Conversely, the amount of self-occlusion decreases when the right foot is moving ahead and therefore the lower section would contain two single connected components which refer to the two feet.

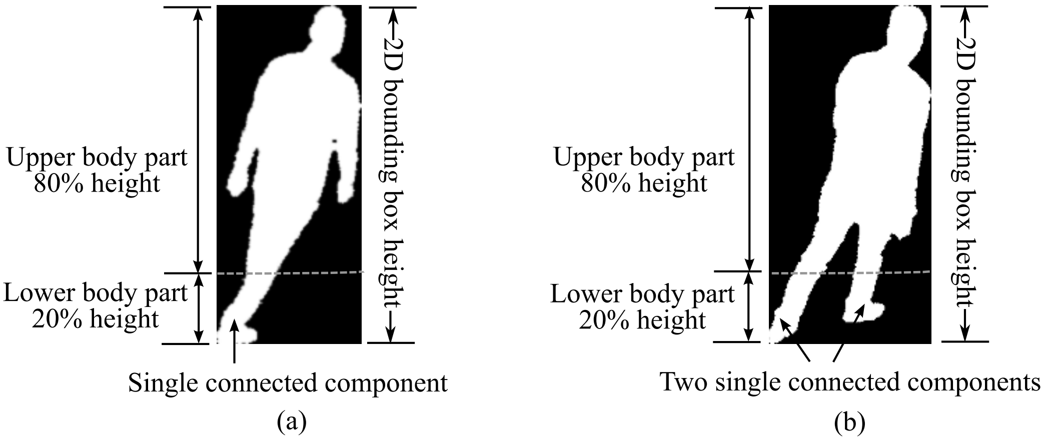


Figure 5.4: Detecting the leading foot in a 2D silhouette: (a) when the left foot is in front, (b) when the right foot is in front.

The visual inspection of the extracted gait cycles and the detecting leading foot showed that it was difficult to determine an accurate gait cycle and leading foot for all gait sequences. As the the detection of a gait cycle and leading foot is not the primary focus of this thesis, we therefore manually corrected them where it was required. A study in [92] also showed how a gait cycle estimation sometimes fails for different viewpoints when its signal had small changes across the walking sequence as it was difficult to detect peaks and valleys.

In this thesis, two gait cycles were extracted from each 2D sequence; one begins when the right foot is in front (right gait cycle), and the other when the left foot is in front (left gait cycle). During the matching process, when the left 3D gait cycle was enrolled it was matched with the left 2D gait cycle and when the right 3D cycle was enrolled it was matched with the right 2D cycle.

5.1.2 3D Alignment and projection

Consider a perspective camera, the appearance of a walking subject varies as he/she moves away from or towards the camera. In order to make the appearance of the synthetic silhouettes generated from 3D volumes similar to that of the real silhouettes from the camera, the 3D volumes in one gait cycle should be aligned before projecting them onto the the image plane of the recorded camera. To compute the 3D alignment, we assumed that the subject is walking on a flat floor along a straight line, which is parallel to the z-axis as shown in Fig. 5.5, and that the camera has been calibrated such that its projection matrix (M) is known. The 2D and 3D gait cycles are constrained so that their phases match (i.e. the first and last frames have the same phase). For each pair of 3D and 2D gait cycle, the first volume and the first silhouette are required to compute the alignment as follows:

Step1: The centroid of the first volume, denoted (x_c, y_c, z_c) is calculated as the position of the 3D gait cycle.

Step2: The walking plane is defined as the position of the centroid along the x-axis (x_c).

Step3: The projection matrix (M) is used to back-project the first silhouette in the 2D gait cycle onto the walking plane (x_c).

Step4: The centroid of the back-projected silhouette is calculated as $(\hat{x_c}, \hat{y_c}, \hat{z_c})$ and is considered as the position of the back-projected 2D gait cycle.

Step5: The amount of displacement (Δz) (in voxels) between the back-projected 2D gait cycle and 3D gait cycle along the z-axis is determined as

$$\Delta z = \hat{z}_c - z_c \tag{5.3}$$

Step6: Each 3D volume, V_k (k is the index of the frame in a 3D gait cycle), in one gait cycle is translated along the z-axis by Δz

$$\grave{V}_k(x, y, z) = V_k(x, y, z + \Delta z) \tag{5.4}$$

Step7: Each aligned 3D volume ($\grave{V}_k$) in one gait cycle is projected onto a 2D image plane of the camera using the general perspective Eq. 3.3 to produce synthetic silhouettes.

It should be noted that the projection process in **Step7** is computed by using a lookup table to pre-compute the mapping between the image coordinates and the world coordinates and speed up the projection process. After that, a frame-by-frame comparison can be computed between the synthetic silhouettes and the real silhouettes from the camera.

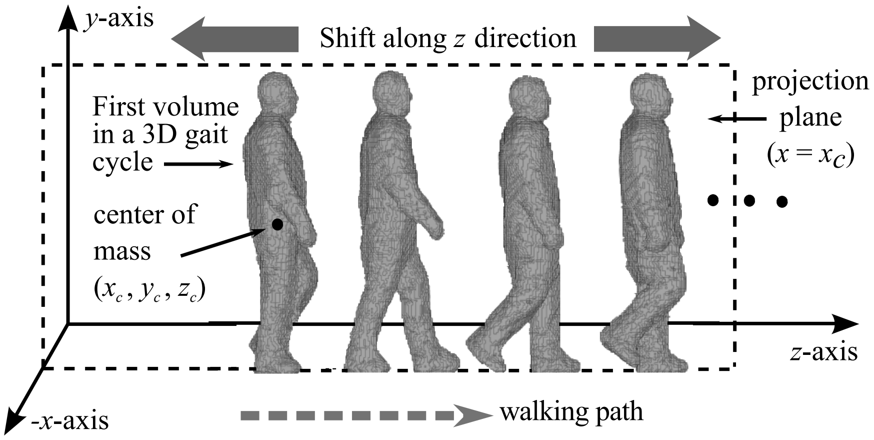


Figure 5.5: Shift a 3D gait cycle along the z-axis.

5.2 Gait feature extraction

The pose of the real and synthetic silhouettes varies from frame to frame. Therefore, the gait features will be extracted on a frame-by-frame basis. We computed Generic Fourier Descriptors (GFDs) from each real and synthetic silhouette over one gait cycle as gait features since GFDs are invariant to many geometric distortions (e.g. translation, scale and rotation) including a small amount of perspective distortion, and have some intrinsic resistance to noise [124]. These invariant properties can be useful to compensate for small transformation errors in the 3D projection process. The GFDs can capture multi-resolution fine features in both radial and circular directions, and require only a small number of features to efficiently describe the shape. To derive the GFD, first the image $f[u,v]$ is represented in polar coordinates $F[r,\theta]$ where $[u,v] \equiv [r\sin\theta, r\cos\theta]$, u and v are the Cartesian coordinates and r and θ are the polar coordinates of the image. To account for translation invariance, the centroid of the silhouette is set as the origin of the polar space. The modified Polar Fourier Transform (G) is then applied as [124]

$$G(\rho,\phi) = \sum_{r}\sum_{i} F(r,\theta_i) \times \exp\left[-j2\pi\left(\frac{r}{H}\rho + \frac{2\pi i}{O}\phi\right)\right] \tag{5.5}$$

where $\theta_i = i(2\pi/O)$, $0 \leq \rho < H$, $0 \leq \phi < O$. H and O are the resolutions of the radial frequency and angular frequency respectively. To make the transform coefficients invariant to rotation and scale, the magnitude of the coefficients is calculated such that the magnitude of the first coefficient is normalized by the polar image area and the magnitudes of the remaining coefficients are normalised by the magnitude of the first coefficient. To improve the computational speed, All the real and synthetic silhouettes are centred and cropped to a fixed size by placing a smallest bounding box that covers the largest silhouette in the dataset. Then, only the first 4 radial and 15 angular coefficients are computed from each cropped silhouette as gait features.

5.3 Dynamic time warping

If the speed of walking changes from one gait cycle to another for the same person or between different people, the resulting feature vectors would be of different lengths. To compare two feature vectors of different lengths, we used Dynamic Time Warping (DTW). DTW [96] is a non-linear time normalisation technique that is used to find the optimal alignment between two sequences by stretching or compressing one of the sequences along its time-axis to match the other. It is commonly used in automatic speech recognition to compare different speech patterns. It has also applications in data mining, robotics, gesture recognition and medicine. Several gait recognition techniques used DTW to effectively match gait patterns of different lengths [112, 13, 3].

To clarify the principles of the DTW algorithm, given two time-dependent sequences $S = (s_1, s_2, ..., s_{M_1})$ and $Q = (q_1, q_2, ..., q_{M_2})$ where M_1 and M_2 are the lengths of S and Q respectively. The objective of the DTW is to compare S and Q. To do that, a local distance matrix E is built to evaluate the local cost measure between each pair of elements of the two sequences. Once this matrix is built, an alignment (warping) path can be traced to extract the corresponding elements in the two sequences. An example of a warping path is shown in Fig. 5.6.

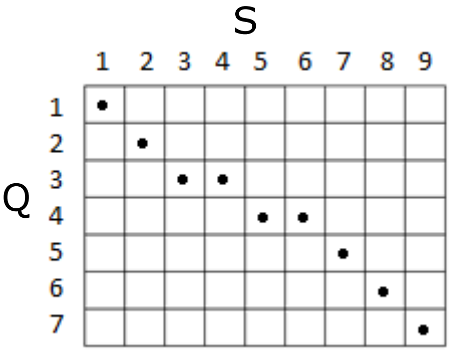


Figure 5.6: Example of a warping path between a sequence S of length M_1 = 9 and sequence Q of length M_2 = 7.

A warping path is represented by a sequence of points $W = (w_1, w_2, ..., w_L)$ with $w(l) = (i(l), j(l))$ where L is the length of the warping path and $i(l)$ and $j(l)$ represent the entries for the first and second sequences respectively. The path must start at the beginning of each sequence and finish at the end of both sequences. This ensures the inclusion of all elements of the sequences in the path. The elements of the sequences need to be matched in a monotonically increasing order to preserve the temporal order of the sequences. Furthermore, all the elements in the path must be pairwise distinct without any duplications. Several valid warping paths can exist with these conditions. To find the best alignment between the comparing sequences, the optimal warping path with a minimal accumulated cost should be selected. In order to extract the best warping path, the DTW algorithm uses dynamic programming to build the accumulated cost matrix D as follows:

$$D(i,j) = \begin{cases} \sum_{k=1}^{j} E(s_1, q_k), & \text{if } i = 1 \\ \sum_{k=1}^{i} E(s_k, q_1), & \text{if } j = 1 \\ \min\{D(i-1,j-1), D(i-1,j), D(i,j-1)\} + E(s_i, q_j), & \text{otherwise} \end{cases} \tag{5.6}$$

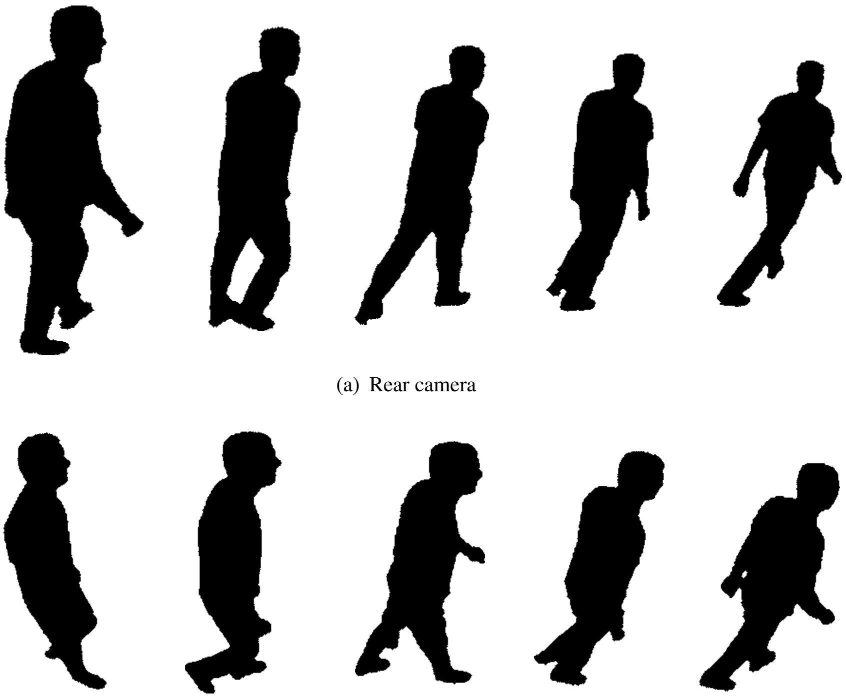
The last element in this matrix represents the minimum accumulated distance between the two sequences. Starting from this element, we can backtrack to extract the best alignment path. In this research, we calculated the Euclidean distance between each pair of feature vectors to compute the local distance matrix E and then applied the DTW algorithm to compute the final accumulated distance.

5.4 Description of a small Soton 3D-2D gait dataset

For evaluation purpose, a small 3D-2D gait dataset was collected in two sessions using the Biometrics Tunnel as a recording site to simulate walking in a narrow corridor as explained in section 4.4. The twelve synchronous cameras in the Tunnel were used to capture people gait images from different viewpoints to reconstruct 3D volumetric data for the gallery set. Two independent cameras were used to record perspectively distorted gait images from the side and rear to build the probe set. The dataset consists of 17 subjects (4 women and 13 men). Each subject walked ten times along a straight line in the middle of the Tunnel in one direction. A total of 170 synchronised multi-view gait sequences and 170 single-view sequences from the rear cameras were recorded. The visual inspection of the gait data from the side camera revealed that three subjects did not have ten walking sequences due to human errors during the capturing process. Therefore, it was decided to consider only 140 sequences, which belong to 14 subjects, from this camera in the analysis. All the subjects were PhD students in the School of Electronics and Computer Science at the University of Southampton. The students' ages ranged between 23 and 30 years old and they belonged to different ethnic backgrounds. Fig. 5.7 shows several frames in one gait sequence from the two arbitrary cameras where the orientation and observation angle change from frame to frame.

5.5 Evaluation results

To evaluate the recognition performance of the proposed technique, the first 60 GFDs were extracted from each synthetic silhouette and real silhouette over one gait cycle. These features were then compared using DTW and Euclidean distance measure. The subjects in all experiments in this thesis were identified as follows: for each probe sequence, we deleted the corresponding 3D volumetric sequence that was captured at the same time from the gallery. Figs. 5.8 and 5.9 show examples of real images captured by the independent rear and side cameras in one gait



(a) Rear camera

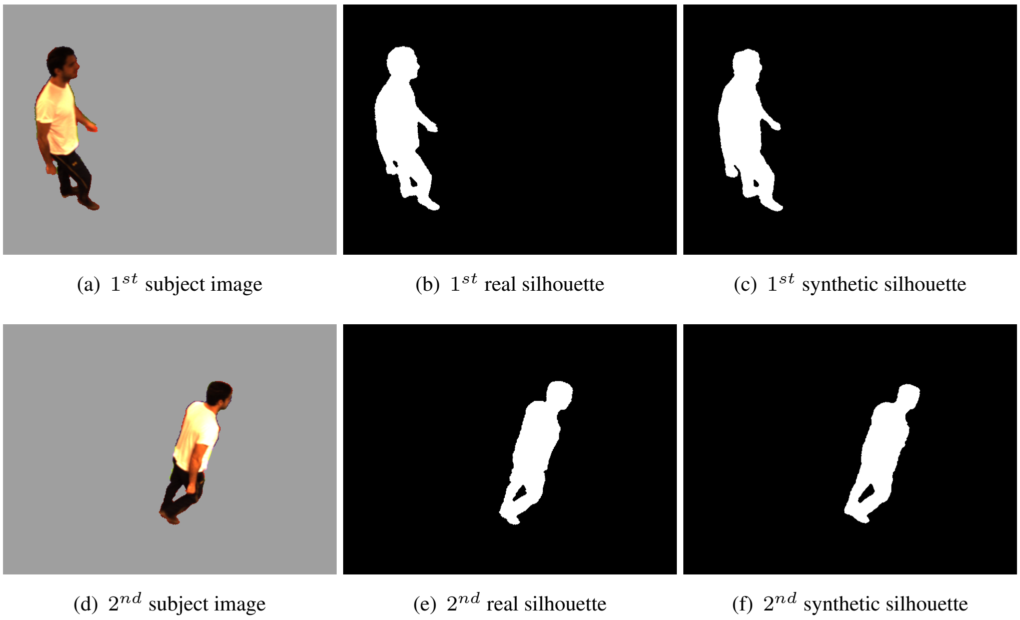
(b) Side camera

Figure 5.7: Several samples of one gait sequence from the (a) rear and (b) side camera for the same person.

cycle and their synthetic silhouettes produced by the proposed technique. In these figures, there is an obvious change in the observation angle in the two frames captured by each of these cameras. In general, the appearance and position of the synthetic silhouettes are similar to those of the subject silhouettes. However, few transformation errors exist due to the 3D reconstruction and projection process.

To measure the identification performance of the proposed technique, we computed the Cumulative Match Score (CMS) [94] which is the percentage of correctly classified test (probe) samples in the top K closest matches of the reference (gallery) samples. The Receiver Operator Characteristics (ROC) curves are also plotted for the main results. Figs. 5.10 and 5.11 show the CMS and ROC respectively for the two cameras using the proposed technique. The Correct Classification Rate (CCR) at rank 1 for the rear and side cameras is 98.8% and 76.4% respectively. In term of ROC, a verification rate of over 90% and 80% is obtained from the rear and side cameras respectively at 20% false positive rate.

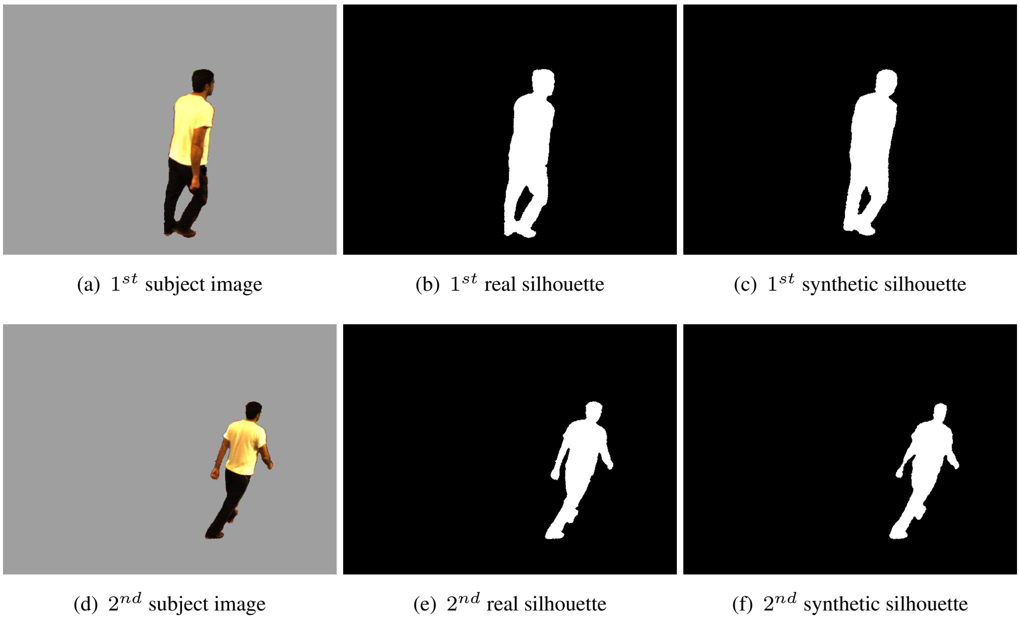
We also carried out the 3-fold cross-validation experiment by dividing the whole dataset into three smaller subsets. In each subset, we used one-third of the silhouette sequences for each subject in the probe set and two-thirds of the 3D volumetric sequences in the gallery set. The average CCRs were $97.4\% \mp 0.9$ and $77.0\% \mp 3.0$ for the rear and side cameras respectively. We noticed that the 3-fold CCR of the rear camera is lower than that using the whole dataset. This



(a) 1^{st} subject image (b) 1^{st} real silhouette (c) 1^{st} synthetic silhouette

(d) 2^{nd} subject image (e) 2^{nd} real silhouette (f) 2^{nd} synthetic silhouette

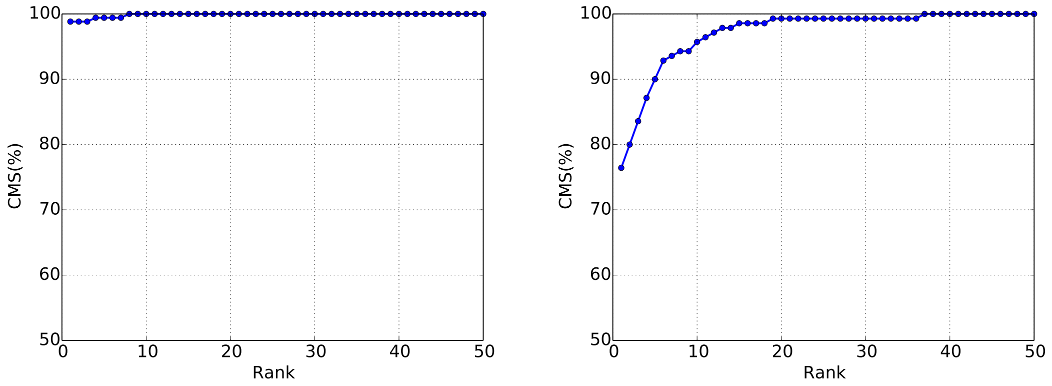
Figure 5.8: Example of real images of a subject in one gait cycle captured by the side camera and their synthetic images produced by the proposed method.



(a) 1^{st} subject image (b) 1^{st} real silhouette (c) 1^{st} synthetic silhouette

(d) 2^{nd} subject image (e) 2^{nd} real silhouette (f) 2^{nd} synthetic silhouette

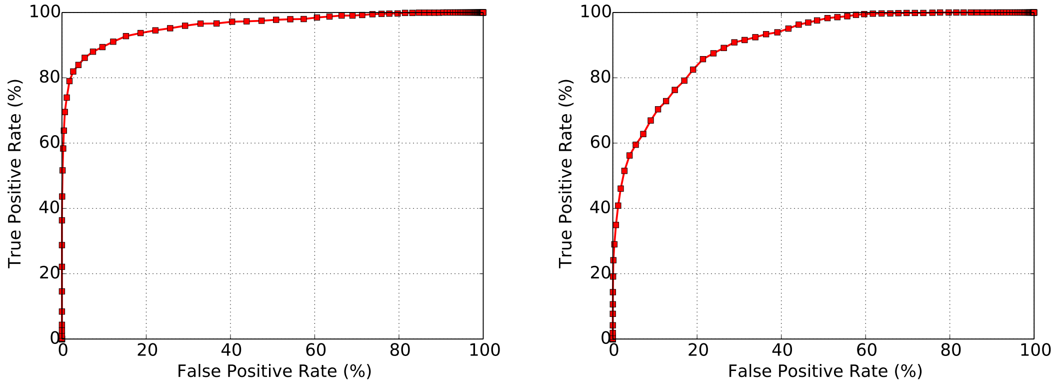
Figure 5.9: Example of real images of a subject in one gait cycle captured by the rear camera and their synthetic images produced by the proposed method.



(a) Rear

(b) Side

Figure 5.10: CMS for the (a) rear and (b) side cameras.



(a) Rear

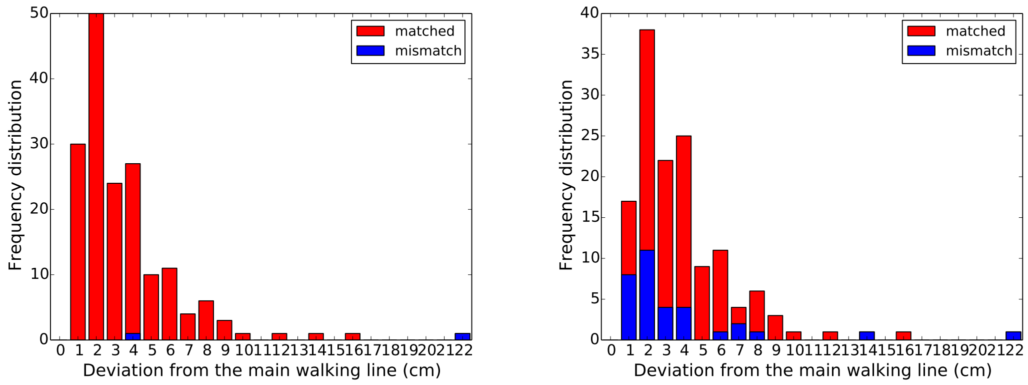
(b) Side

Figure 5.11: ROC curves for the (a) rear and (b) side cameras.

is because the CCR depends on the dataset; particularly on the number of sequences for each subject in the gallery set.

The results, in general, take the same trends as those obtained in section 3.3.4 using the GEI but they are slightly better than those from the far cameras and much better than those from the middle cameras. The improvement in the performance is probably due to (1) the enhancement in the quality of the 3D volumetric data when using all the twelve cameras in 3D reconstruction process, (2) 3D alignment and (3) frame-by-frame comparison. However, the recognition performance from the side camera is still low. Further investigations were performed to understand the results.

At the beginning, we visually inspected the two misclassified samples from the rear camera. We noticed that the subjects appeared to deviate from the main walking line. After that, we decided to measure the deviation of people from the main walking line numerically as follows: (1) the centroid values along the x- and z-axis for the first, middle and last volumes in each 3D gait cycle were computed as local walking trajectories, (2) the global walking trajectory was calculated by computing the mean of the centroid points over all 3D gait cycles, (3) a straight line was then fitted into the mean centroid points using the least square method to estimate the main walking line, (4) the mean distance error between each local walking trajectory and the main walking line was then computed to measure the amount of deviation. The histogram for the deviation was plotted in Fig. 5.12. This histogram shows that the sample with the largest mean distance error (largest deviation) from the main walking line was misclassified in the two cameras. The deviation can cause an error when computing the 3D alignment for the volumetric data and subsequently create a discrepancy between the synthetic and the real silhouettes. Furthermore, the histogram indicates that some other samples with smaller distances had also been misclassified. We hypothesised that some shape details could be lost when extracting gait features from the whole silhouette region which could affect the discriminatory power of the features. Therefore, we did an experiment by dividing each silhouette along one gait cycle into multiple areas (i.e. different body parts) and extracting gait features from each area. Several studies [23, 118, 32, 1] showed that fusing features at the score level can boost the overall recognition performance. Therefore, the features from each area were matched separately. The final matching score was computed by fusing the matching results of all areas at the score level using the sum rule. Table 5.1 shows how the recognition rate increases from 76.4% to 89.2% and then to 95% for the side camera when dividing the silhouette into 2 (vertical) parts and then into 4 (2 vertical and 2 horizontal) parts respectively as different body parts could be best described. Further divisions will be analysed in the next chapter.



(a) Rear

(b) Side

Figure 5.12: The relationship between the deviation from the main walking line and misclassified samples from the (a) rear and (b) side cameras.

Table 5.1: Performance versus no. of divisions in a silhouette.

No. of divisions	1	2	4
CCR(%)	76.4	89.2	95.0

The processing times required to do the key operations in this technique are listed in Table 5.2. All the computations were done on an Intel computer with a Core 7 CPU 3.40GHz and RAM 16GB. The time required to do 3D alignment/projection process for one volume is approximately four times longer than that required to extract 60 GFDs from one image. However, the projection time could be improved by a factor of two when optimising the code and reducing the complexity of the inner loops. The processing time could be further speeded up by exploiting the capabilities of commonly available graphics cards [55].

Table 5.2: Processing time for the key operations.

Operation	3D alignment/projection	60 GFDs
Avg. time(sec)/frame	0.4	0.1

5.6 Performance analysis

This section demonstrates the effectiveness of the proposed technique using a number of experiments. In these experiments, we used the silhouettes of all subjects from the rear camera for the probe set and the 3D volumetric data for the gallery set. These experiments include the influence of using different GFD resolutions, noisy silhouettes in the probe, different proportions of the gait cycle length, perspective projection sensitivity from different viewpoints and camera calibration errors. The estimated error in the experiments was calculated when the number of corrected classified samples (sequences) changes by one unless it states otherwise.

5.6.1 GFD resolution versus performance

The resolution of the GFD can be determined by the number of radial and angular (orientation) features extracted. Zhang and Liu [124] used GFDs for shape representation and retrieval. They evaluated the performance of GFDs using different numbers of features and found that a small number of features was sufficient to achieve high retrieval accuracy. Here, we used GFDs for gait recognition and it is crucial to select an optimal number of GFD features. To that end, we had evaluated CCR over a range of different numbers of radial and angular features: 3 to 5 radial features and 6 to 20 angular features. Fig. 5.13 shows the relationship between the recognition performance and GFD resolutions.

As can be seen that high CCRs can be achieved using many different combinations of features using Eq. 5.5. Even with a low number of features, 18 per frame, the recognition rate is still above 90%. Furthermore, the performance is improved by increasing the number of angular features (O). Meanwhile, the performance is not enhanced by increasing the number of radial features (H) for a constant angular resolution. This suggests that angular features have more influence on the recognition performance than radial features. The results also indicate that a suitable resolution to achieve an effective recognition performance with the smallest number of

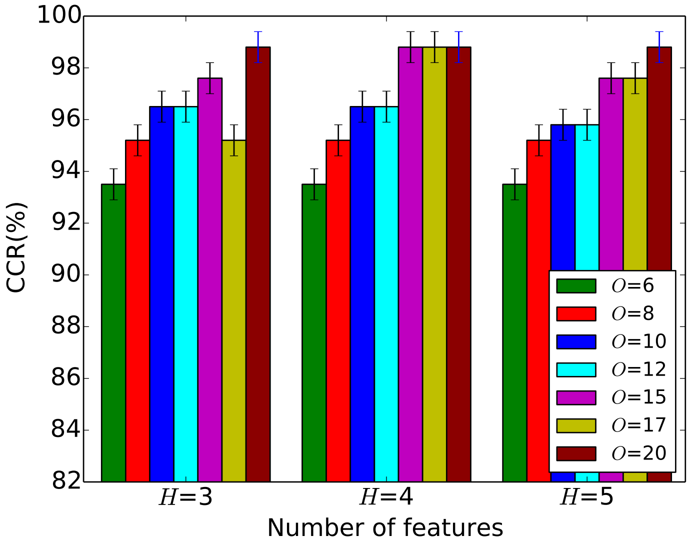


Figure 5.13: Number of features versus performance where H refers to the number of radial features and O refers to the number of angular features.

features is obtained by using 4 radial features and 15 angular features (60 GFDs). This resolution will be used for feature extraction in the remaining analyses.

5.6.2 The Effect of noise

The measurements were made in a constrained environment with good lighting and digital recording. Therefore, it is important to investigate the sensitivity of the technique to noisy silhouettes. An experiment was conducted to test this where salt and pepper noise was added at different rates (from 10% to 40%) to the probe silhouettes, while the synthetic silhouettes in the gallery remained unchanged. For each noise rate, the experiment was repeated three times and the average CCR and standard deviation were calculated. The evaluation results are shown in Fig. 5.14, where the performance slightly degrades as the noise rate increases. The averaged CCR is above 95% for different noise rates. The maximum degradation in performance is only around 2% for 40% added noise. These findings indicate that the technique is insensitive to this type of noise. This may be due to considering the whole shape via the use of the Fourier Transform in the feature extraction process.

5.6.3 The Influence of a truncated gait cycle

Little information in the literature is available about how the recognition performance can be influenced by reducing the number of frames in one gait cycle and which part of the cycle contains more discriminative information. The influence of using a truncated gait cycle was measured and the recognition performance in different parts of a gait cycle was evaluated. To do that,

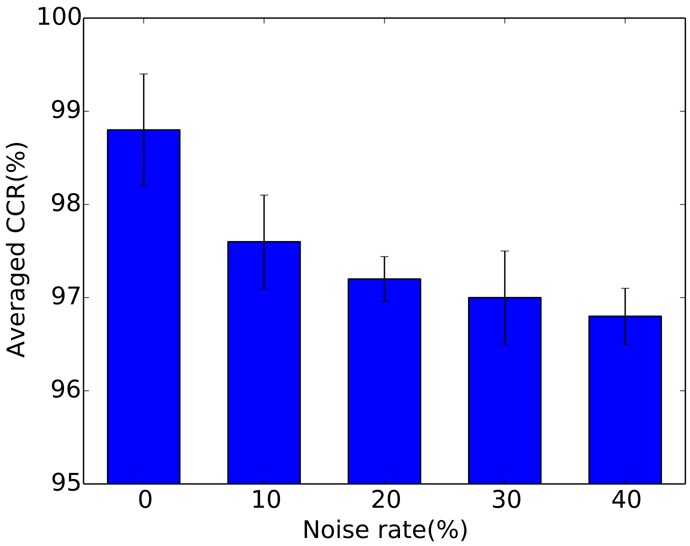


Figure 5.14: The impact of using noisy silhouettes.

we considered proportions of 25%, 50% and 75% of the total number of frames in a gait cycle. These proportions were selected from the start, middle and end of each gait cycle to evaluate the effective part of a gait cycle for recognition. The results are illustrated in Fig. 5.15. As can be seen that using a truncated gait cycle slightly degrades the performance where the maximum drop in recognition rate is around 3% when only 25% of the total gait cycle length was used for recognition. In general, the performance is slightly improved by using more frames from the start and end of a gait cycle as there is more information available for recognition. However, there is a fluctuation in performance for the middle of a gait cycle where the highest CCR is obtained when using only 25% of the frames and the lowest is achieved when using 50%. This fluctuation might be attributed to the small number of samples in the gait dataset.

In terms of the position in the gait cycle, the average CCR over different proportions was calculated at the start, middle and end of a gait cycle. The average CCR at the end of a gait cycle (96%) is lower than that of the other parts. This is probably due to the smaller resolution of the silhouettes at this part (the subjects walk away from the camera) which could result in losing fine details and producing less discriminant features. On the other hand, the best CCR is achieved at the middle of a gait cycle (97%). By observing the middle part in all gait cycles, it was found that this part includes the non-occluded motion of the leg compared to the other parts which may lead to the extraction of more discriminant features. These results demonstrate the efficiency of the proposed technique in handling a truncated gait cycle.

5.6.4 Sensitivity of perspective projection distortion

In this experiment, we measured the sensitivity of the perspective distortion from different viewpoints on the recognition performance. Instead of physically installing an arbitrary camera

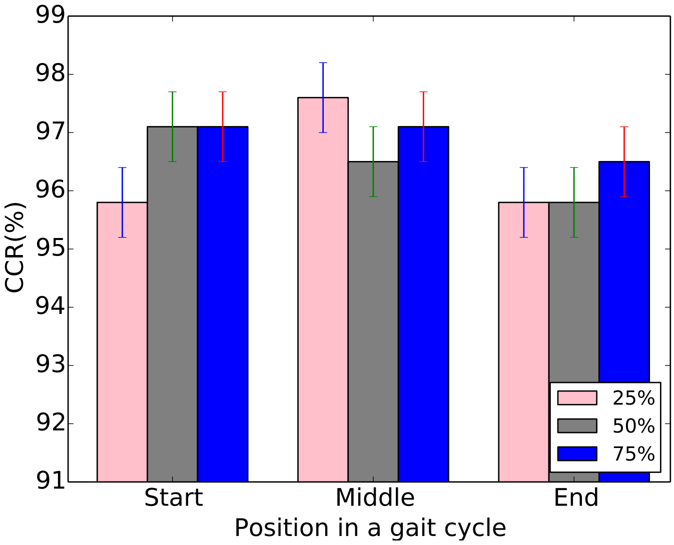


Figure 5.15: Recognition rate vs. truncated gait cycle.

at various positions in the Tunnel for this sensitivity analysis, we used a half number (i.e. five) of 3D sequences for each person in the dataset to generate synthetic silhouette sequences from independent viewpoints for the probe set [see Fig. 5.16] and the remaining 3D sequences for the gallery set. Projection matrices of the arbitrary views were generated from the projection matrix of one of the middle cameras. We decomposed the projection matrix into an intrinsic and extrinsic (position and orientation) matrices. We used the same intrinsic matrix and distortion parameters of that camera and modified the extrinsic matrix by placing a virtual camera at various positions on both sides of the Tunnel. The 3D alignment/projection process was then computed and the subjects were identified using GFDs as gait features.

In order to illustrate the powerful of the GFDs for gait recognition from different viewing angles, we also calculated gait features using (1) the Gait Energy Image (GEI) [72], due to its higher recognition performance and wide use for comparison, and (2) the Affine Moments Invariants (AMIs) [28] because they also have invariant properties under geometrical transformations (rotation, translation and scale) as the GFDs and show high performance for gait recognition [43]. For a fair comparison, we extracted 60 AMIs from each gallery and probe image. The recognition results are shown in Fig. 5.17 for several synthetic views that are different from all the views in the Tunnel. We can see that the GFDs have a superior performance with the best recognition rate of 100%. However, the recognition rate decreases as the optical axis of the virtual camera gets closer to the front/rear view because errors in the projection onto the walking plane increase. Ultimately, the camera axis becomes parallel to the walking plane and the problem is ill-posed. We noticed that the performance of the GEI is better than that obtained using the perspective silhouettes from the middle cameras in chapter 3. This may be attributed to the 3D alignment process which removed the significant variation in the appearance of silhouettes in the gallery and probe sets. However, the GEI performs worse than the GFDs because the relevant body parts could not effectively be matched when computing the average image since the orientation of the silhouettes from the virtual viewpoints changes at each frame. Fig. 5.18 shows an example for the calculated GEIs from some viewpoints. Furthermore, the AMIs do not perform well for the perspective silhouettes in this experiment where a good recognition rate can only be obtained for the side view (i.e. 90°). It should be noted that the matching technique cannot be applied when the optical axis of the probe camera is parallel to the walking direction as we cannot compute the projection onto the walking plane. In general, the results demonstrate that the proposed technique is not sensitive to the perspective distortion from different viewpoints using the GFDs as gait features.

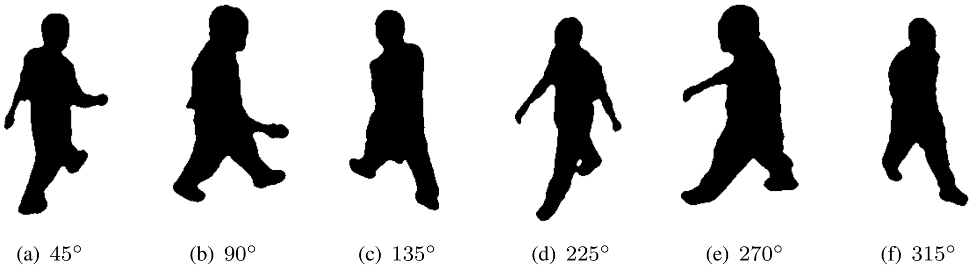

(a) 45° (b) 90° (c) 135° (d) 225° (e) 270° (f) 315°

Figure 5.16: Synthetic images from different viewpoints.

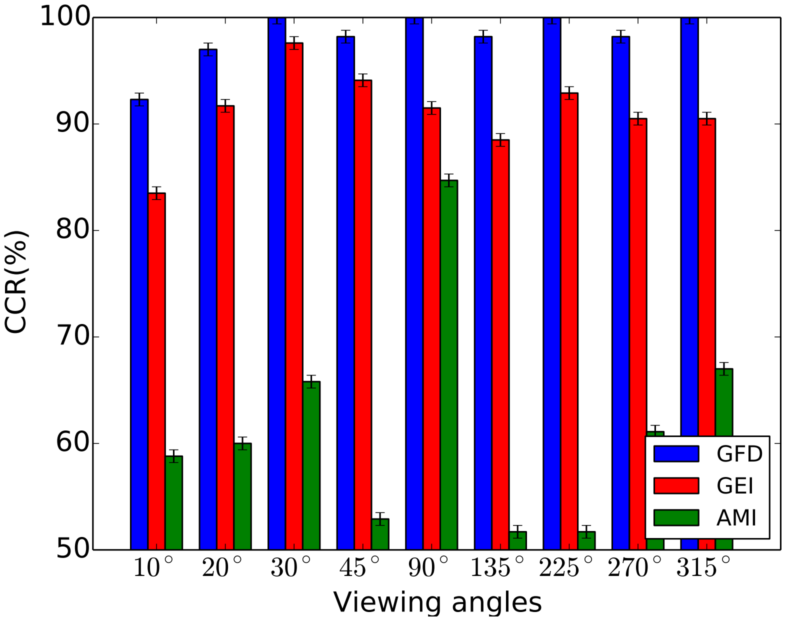


Figure 5.17: Sensitivity of the perspective distortion on performance.

Figure 5.18: GEIs calculated from different virtual viewpoints for the same person.

5.6.5 The Impact of Camera Calibration Accuracy

To compute an effective match between synthetic silhouettes and original silhouettes from a real camera, the synthetic silhouettes should be sufficiently accurate. This would depend on the quality of the calibration for the arbitrary camera at the first place, which in turn affects the recognition performance. Therefore, an analysis was performed to investigate the sensitivity of recognition performance towards calibration errors from an arbitrary camera. A translational error was introduced to the projection (calibration) matrix, and then the gait data in the gallery and probe were processed accordingly. The original calibration points for the camera, which include 2D-3D corresponding points, were used to interpret the change to the calibration matrix. The 3D points were projected onto the image plane using the erroneous calibration matrix and the mean distance error between the original 2D points and the projected 2D points was calculated for different error levels.

Fig. 5.19 shows that the performance significantly degrades as the level of the translational error increases. When the mean distance error is less than 6 pixels, the recognition rate is above 90%. The built in invariance of the chosen features can compensate for small transformation errors in the projection process. However, there is a sharp decrease in recognition rate when the mean distance error becomes greater than 6 pixels. These findings reveal the sensitivity of the technique towards calibration errors, which emphasise the importance of accurately characterising the camera. This is expected as the matching stage mainly depends on the accuracy of the 3D alignment and projection process. Hence, the presence of errors in the calibration matrix could influence the amount of displacement between 2D and 3D gait cycles and consequently the appearance of the synthetic silhouettes, which in turn creates a discrepancy between the synthetic and the real silhouettes.

5.7 Discussion

In this chapter, a new technique was presented to identify subjects walking along a straight line using gait images captured by a perspective camera mounted on the wall in a narrow corridor where the shape of the subject is distorted according to his/her position in recorded images. The technique is mainly based on a 3D alignment/projection process to allow silhouettes in each pair

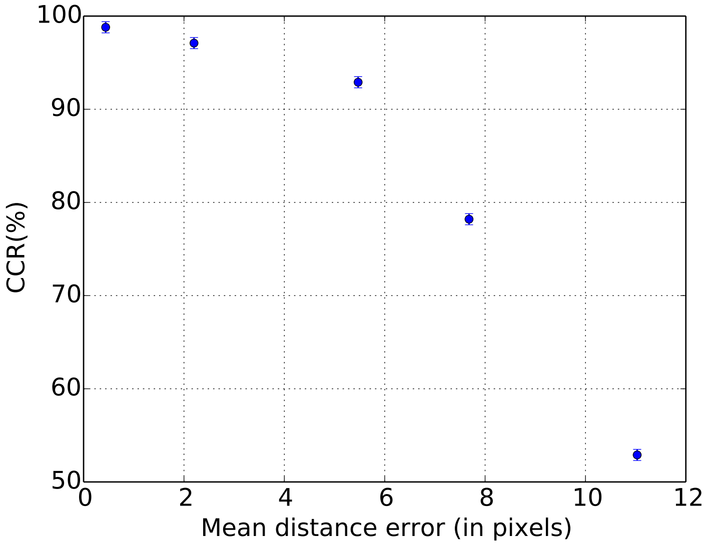


Figure 5.19: The effect of calibration errors upon performance.

of a gallery and probe gait cycle share the same appearance, orientation and observation angle. In order to accurately estimate a gait cycle and leading foot in all gait sequences, several efforts were spent. However, no ideal approach was found because the results are so data-dependent. Since the primary application of the work done in this thesis is in a forensic scenario where we have one or a small number of probe sequences, manual detection was used to accurately label a gait cycle and leading foot. The main aim, in this thesis, is to develop an efficient matching technique.

The proposed technique in this chapter was validated on a small gait dataset collected inside the Biometrics Tunnel where two cameras with rear and side views were used for testing and the twelve synchronous cameras were used for enrolment. The evaluation results showed a high recognition performance by the rear camera and a low performance by the side camera. The misclassification of samples from the two cameras could be attributed to the deviation of people from the main walking line. This led to errors when computing the 3D alignment which consequently created a discrepancy between the synthetic silhouettes from the 3D volumes and the real silhouettes. Further investigation revealed that fusing features from different body parts significantly improved the recognition performance from the side camera since more detailed features could be extracted.

The performance analysis from the rear camera showed that high recognition could be achieved using a small number of GFDs since these features are extracted from the spectral domain which make them more concise than those extracted from the spatial domain. Angular (orientation) features are shown to have more influence on recognition performance than radial features. Our hypothesis is that using more angular features give better handling of the orientation changes of silhouettes in the dataset. Further analysis demonstrated that the technique is insensitive to noisy silhouettes which may perhaps due to the resilience of the selected features.

The effect of using gait cycles truncated in different proportions was evaluated. The results showed the efficacy of the technique in handling truncated gait cycles where the recognition rate only dropped slightly even when only 25% of the total gait cycle length was used for recognition. This means that the gait features extracted from the same part of a gait cycle could provide a sufficient discriminatory capability. It was also illustrated that the recognition performance using frames at the end of a gait cycle was lower than those from other parts. This might be due to the smaller resolution of silhouettes at this part (the subjects walked away from the camera) which could result in the loss of fine details and a reduction in the discriminatory power of the gait features.

Another analysis showed that the technique had a high resilience against perspective distortion from several virtual viewpoints on both sides of the Tunnel using the GFDs as gait features where high recognition rates were recorded from the selected viewpoints. This might be due to using the same type of gait data for both enrolment and recognition. The main limitation of the proposed matching technique is that it is not applicable when the optical axis of the probe camera is parallel to the walking direction because we cannot compute the projection onto the walking plane. This analysis also showed that gait features derived from the GFDs have a higher discriminatory capability than those from the GEIs and AMIs. The sensitivity of the technique towards calibration errors from the probe camera was also measured. The results revealed that the recognition performance significantly deteriorated in the presence of calibration errors. This is expected as aligning 3D volumetric data and producing synthetic silhouettes according to the probe camera viewpoint mainly depend on the quality of its calibration. Consequently, the presence of calibration errors will lead to variations in the appearances of the synthetic and real silhouettes.

Based on this discussion, the following points are considered for future investigation in the next chapter: (1) the GFDs will be used as gait features because of their higher discriminatory power, (2) an another matching technique, which is completely view-invariant, should be developed to tackle the limitations of the current technique and (3) a larger gait dataset should be collected to better estimate the inter-class variation between subjects.

Chapter 6

Gait Recognition for Straight and Curved Trajectories

The results of the analyses in chapter 5 highlight several limitations in the previous proposed matching technique (algorithm). These limitations are explained in the following points:

- It was assumed that people walked only along a straight line. According to that the starting position of a gait cycle was only considered to compute the 3D alignment. However, this assumption is not always true as people normally deviate to the left or right even if they are walking along a straight line.
- The previous algorithm used the walking plane of the enrolled people in the gallery to estimate the position of the back-projected 2D gait cycle in 3D space. This made the algorithm inapplicable when the optical axis of the camera is parallel to the walking direction (i.e. camera views the subject from a front/rear view).

In this chapter, we developed a new matching algorithm to cope with the previous limitations. The 3D alignment in the new algorithm will be computed based on the three key frames at the start, middle and end of a gait cycle. The direction of walking between these key frames will also be considered. Therefore, the algorithm will not be limited to walking only along a straight line. To make the algorithm completely viewpoint invariant, a new procedure is proposed to estimate the position of the subject in the three key frames of a gait cycle. Furthermore, it was shown in the previous chapter that the discriminatory power of gait features derived from the global body shape using the Generic Fourier Descriptors (GFDs) can be improved by extracting more detailed features from different body parts. Therefore, it is decided in this algorithm to divide the silhouette into several areas and to extract the features from each area for recognition. Finally, the small gait dataset used in the analyses in chapter 5 is not large enough to accurately estimate the intra-class variations between subjects. A new larger dataset was collected as described in chapter 4 to evaluate the performance of the new algorithm in this chapter. This dataset includes people walking along straight and curved trajectories to investigate how the recognition performance changes for a continuously changing direction in one gait cycle. More details about the new algorithm and the larger gait dataset will be given in this chapter.

6.1 The revised algorithm

A revised version of the previously proposed algorithm in chapter 5 was developed to identify people regardless of their walking direction and independently of the captured camera viewpoint. The revised algorithm uses 3D volumes of the people in the gallery and an alignment procedure which requires only the position of the three key frames in each gait cycle. It starts with the detection of these frames at the start, middle and end of a gait cycle. They are then used to divide the gait cycle into two parts. The position of a subject in the key frames is estimated in 3D space. Here, we assume that the motion in the first and second part of a gait cycle can be approximated as linear even if the walking trajectory is a curve. Accordingly, the position of the intermediate frames and the local walking direction are calculated in the first and second part of a gait cycle. Each subject image is then compared with the corresponding synthetic image generated from the aligned 3D volume using Generic Fourier Descriptors as gait features. The key operations in this algorithm will be explained in the following subsections.

6.1.1 Key frame selection

During walking, the body of the subject will swing to the left or right depending on which foot is moving forward even if he/she walks along a straight line. The swing will reach a maximum when the two feet are at full extension (double-support stance). On the other hand, when the two feet are closest to each other (mid-stance), the swing towards left or right will be at the minimum. Three consecutive frames at mid-stance will be chosen as the key frames to mark the start, middle and end of a gait cycle.

6.1.2 Estimating positions and walking directions

In order to adapt to view variation due to walking direction changes, the position and walking direction (pose information) of the (a) 2D subjects and (b) 3D volumes of people are estimated as follows:

6.1.2.1 Estimate the pose information from the 2D silhouettes of the subjects

We assume that a subject walks along any direction on a flat floor and that the recorded camera is calibrated such that the camera projection matrix (M) and the location of the floor in 3D space are known. The position of the subject in 3D space is estimated by back-projecting his/her foot points onto the floor when the feet meet the floor. Iwashita, et al. [44] estimated the position of the foot by applying the PCA to a silhouette region and finding a convex hull of the silhouette region. Then, the lower intersection point from the eigen-vector of the first principal component and the convex hull is extracted and back-projected onto the floor to estimate the foot position of the subject in 3D space. However, estimating the foot position from only a single point may not be reliable due to imperfections in silhouette extraction. Therefore, we proposed the following procedure to estimate the foot position using most of the points in the foot region:

1. The boundary points of the lower (11%) silhouette region are extracted and their profile is smoothed,
2. The local minima (Q_i) of the lower boundary points are determined,
3. The points that lie within a small vertical distance to Q_i, we set the distance to be 5 pixels, are located as shown in Fig. 6.1,
4. The camera projection matrix (M) is used to back-project the extracted points onto the floor plane (Y = 1) to calculate their positions in 3D space.
5. The mean (f) of the back-projected foot points is calculated as the foot position of the subject in 3D space

$$f = \frac{\sum_k f_k}{K} \tag{6.1}$$

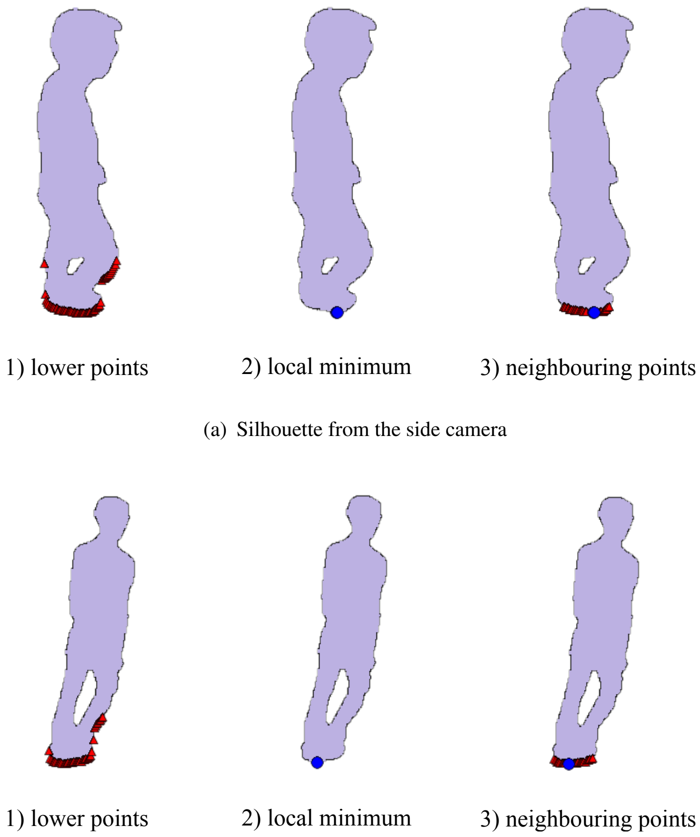
where K is the number of the back-projected foot points, $0 \leq k < K$ and f_k is the back-projected foot point in 3D space. Using this procedure, the foot position of the subject in the three key frames at the start, middle and end of a 2D gait cycle are estimated.

As the motion in the first and second part of the gait cycle is assumed to be linear, the foot positions of the subject in the intermediate frames are determined by fitting two line segments in the two parts of the gait cycle using linear interpolation as

$$\begin{aligned} f_{1i} &= f_s + i \times \frac{f_m - f_s}{L_1 - 1} \qquad i = 0, 1...L_1 - 1 \\ f_{2i} &= f_m + i \times \frac{f_e - f_m}{L_2 - 1} \qquad i = 0, 1...L_2 - 1 \end{aligned} \tag{6.2}$$

where f_s, f_m and f_e are the foot positions at the start, middle and last frame of the gait cycle. L_1 and L_2 are the number of frames in the first and second part of the gait cycle while f_{1i} and f_{2i} are the foot positions in 3D space at a frame i in the first and second part respectively. After that, the local walking directions (φ_1) and (φ_2) between the first and middle key frames and between the middle and last key frames of the gait cycle respectively are calculated from the gradient of the two fitted line segments $\overrightarrow{d_1} = f_m - f_s$ and $\overrightarrow{d_2} = f_e - f_m$ respectively as

$$\varphi_1 = arctan\frac{d_{1_x}}{d_{1_z}} \qquad \varphi_2 = arctan\frac{d_{2_x}}{d_{2_z}} \tag{6.3}$$



(a) Silhouette from the side camera

(b) Silhouette from the rear camera

Figure 6.1: Foot points detection in a 2D image.

6.1.2.2 Estimate the pose information from the 3D volumes of people

To calculate the pose of each 3D volume along one gait cycle, the foot positions of the three key volumes at the start, middle and end of a 3D gait cycle are first computed and the local walking directions between these key frames are then estimated. To determine the foot position of each key volume, the lower (11%) part of the 3D volume is mapped onto the floor using orthogonal top projection to extract the foot region. This step aims to remove the effect of the upper limb movement when extracting a foot position. After that, the middle point (i.e. the centroid) of the foot region is calculated as a foot position of the volume. The foot positions of the intermediate volumes in the first and second part of the gait cycle are computed by fitting two line segments using linear interpolation as

$$\begin{aligned} F_{1i} &= F_s + i \times \frac{F_m - F_s}{N_1 - 1} \qquad i = 0, 1 \ldots N_1 - 1 \\ F_{2i} &= F_m + i \times \frac{F_e - F_m}{N_2 - 1} \qquad i = 0, 1 \ldots N_2 - 1 \end{aligned} \tag{6.4}$$

where F_{1i} and F_{2i} are the foot positions of the volumes in the first and second part of the 3D gait cycle respectively, F_s, F_m and F_e are the foot positions of the first, middle and last volume in the

3D gait cycle. N_1 and N_2 are the number of the volumes in the first and second part of the gait cycle. Next, the gradient of the two fitted lines is used to calculate the local walking directions (δ_1) and (δ_2) between the the first and middle key frames and between the middle and last key frames respectively as

$$\delta_1 = arctan\frac{F_{m_x} - F_{s_x}}{F_{m_z} - F_{s_z}} \qquad \delta_2 = arctan\frac{F_{e_x} - F_{m_x}}{F_{e_z} - F_{m_z}} \tag{6.5}$$

6.1.3 3D alignment and synthetic image generation

In order to align the position and orientation of each 3D volume in one gait cycle with those of the corresponding subject images, we normalised the number of frames in all gait cycles such that the phases in both 3D and 2D gait cycles are matched. More specifically, we normalised the number of frames in the first and second part of a 3D gait cycle using $\acute{i_1} = i_1 \times N_1/\acute{N_1}$ and $\acute{i_2} = i_2 \times N_2/\acute{N_2}$ where $\acute{N_1}$ and $\acute{N_2}$ are the normalised number of frames in the first and second part respectively. The number of frames in a 2D gait cycle is normalised in the same manner. To align the orientation, a 3D coordinate system is set up at each foot position and then a 3D volume is rotated around the vertical axis (y-axis). The volumes in the first part of the 3D gait cycle are rotated by $(\varphi_1 - \delta_1)$ and in the second part by $(\varphi_2 - \delta_2)$. To align the position, each volume in the first part of the 3D gait cycle is translated by ΔF_{1_i} and in the second part by ΔF_{2_i} as

$$\begin{aligned} \Delta F_{1_i} &= f_{1_i} - F_{1_i} \\ \Delta F_{2_i} &= f_{2_i} - F_{2_i} \end{aligned} \tag{6.6}$$

Finally, each aligned volume is projected onto a 2D image plane using the general perspective Eq. 3.3 to produce a synthetic silhouette. This is the most computationally intense part of the algorithm and it was carried out by using a lookup table.

6.1.4 Gait features extraction

The Generic Fourier Descriptors (GFDs) are calculated as gait features because of their invariant properties and higher discriminatory power as illustrated in chapter 5. The invariance of these descriptors to many geometrical distortions (rotation, translation, scale and a small amount of perspective distortion) are exploited to tackle errors that may result in foot position estimation. Section 5.2 describes how to derive the GFDs from the image. It has been shown that the discriminatory power of the GFDs is improved by extracting detailed features from different body parts. Therefore, it is decided to crop the silhouette region such that the centroid of the silhouette is at the middle of the cropped silhouette image. After that, the cropped image is divided into multiple areas of equal size as shown in Fig. 6.2. Then, the first 60 GFDs (4 radial and 15 angular features) are calculated from each area and matched separately. The results of matching gait features from each area are fused at the score level using the sum rule.

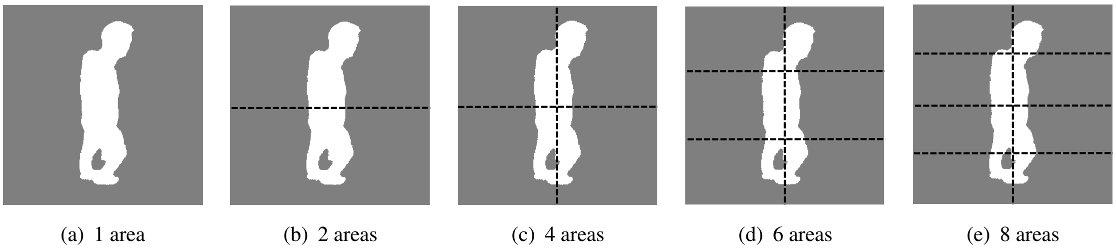


Figure 6.2: The divided areas in a silhouette.

6.1.5 Gait classification

In the previous matching algorithm 5.1, all the frames in a gait cycle were used for matching. Therefore, the resulting gait feature vectors were of different lengths. In order to match gait features of the gallery and probe, we used the DTW algorithm as described in section 5.3. In the revised version, we normalised the number of frames in each gait cycle in order to compute the 3D alignment so that all gait cycles have the same length and a frame-by-frame comparison can then be computed directly. The normalisation process and DTW algorithm do the same job. Furthermore, the recognition results in sections 5.5 and 6.3.1 using the DTW and normalisation process respectively are similar. This means that no significant advantage is achieved using the DTW. Therefore, we choose the simpler approach (i.e. normalisation) to compute the recognition results in this chapter.

However, a frame-by-frame comparison can be affected by errors in gait cycle estimation. To reduce the effect of these errors, we propose the following procedure to compute the distance between two gait cycles. The features of each frame in the probe gait cycle are compared against the features of its corresponding frame, the frame before and the frame after in the gallery cycle. We then take the minimum, D_i

$$D_i = \min(||p_i - g_{i+j}||),\ j \in -1, 0, 1 \tag{6.7}$$

where p and g are the feature vectors of the probe and gallery gait cycle respectively and i and $i + j$ refer to specific frames. At the ends of the cycle, only 2 comparisons will be made. Then, the total distance between two gait cycles of length Z is computed as

$$Dist(p, g) = \sum_{i=0}^{Z-1} D_i \tag{6.8}$$

6.2 Gait datasets

To evaluate the performance of this technique for gait recognition on straight and curved trajectories, two gait datasets that represent different situations were used. The first dataset (large

Soton 3D-2D dataset) was collected in eight sessions inside the Biometrics Tunnel simulating a narrow walkway in airports (access control). Fifty people participated where each one walked six times along a straight line in the middle of the Tunnel and six times along a curved trajectory as indicated in Fig. 6.3. The curved trajectory composed of a straight line at the beginning, then a turn around the middle and finally a straight line. The twelve synchronous cameras were used to capture people from 12 different views to build their 3D volumetric reconstructions for the gallery set. The probe set consists of gait images captured from two independent views by the asynchronous rear and side cameras. All the cameras had the same settings as those used to collect the smaller gait dataset in chapter 5. The experimental setup used to collect this dataset was explained in section 4.4.

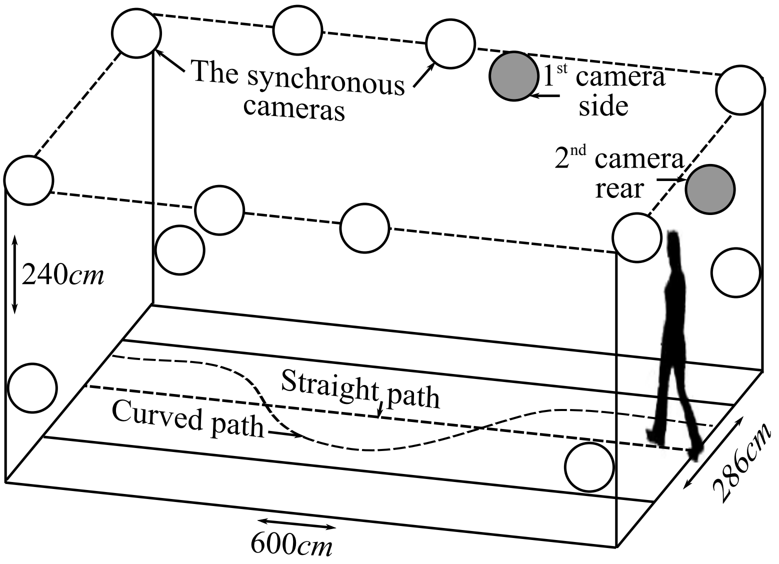


Figure 6.3: Camera positions and trajectories in the large Soton 3D-2D gait dataset.

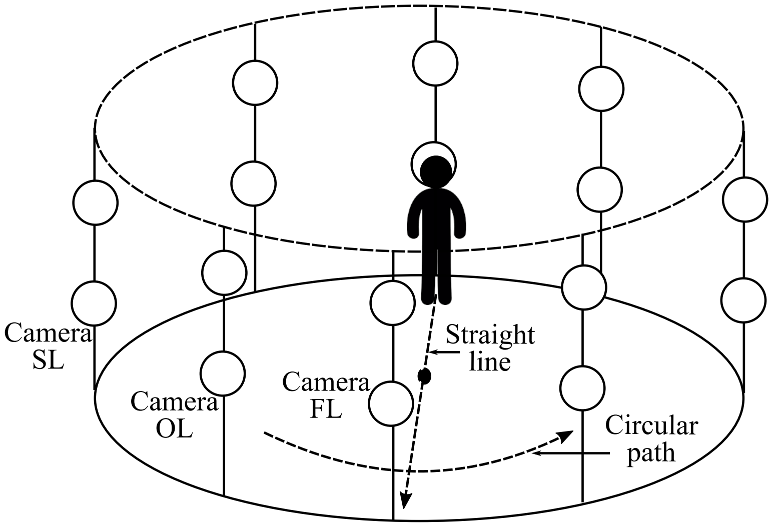
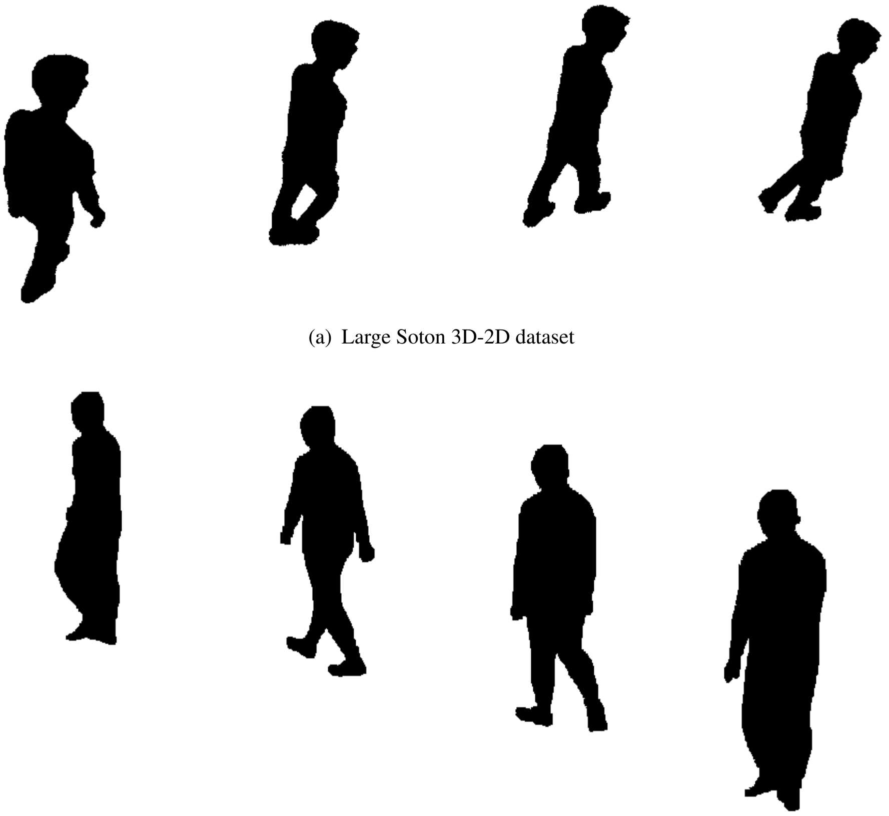


Figure 6.4: Circular studio for the KY4D dataset.

The second dataset is the Kyushu University 4D Gait Database (KY4D) [44]. A wide circular studio was used to record 42 people by 16 synchronous cameras arranged at two different heights. Each subject walked four times along straight lines and once along each of two circles of radius $3m$ and $1.5m$ as shown in Fig. 6.4. Although this dataset includes 3D volumes reconstructed using the images from all the cameras, we have reconstructed them by excluding the probe (target) camera using the shape from silhouette reconstruction [16]. To test the performance of gait recognition for straight walking, we used gait images captured by each of the three cameras from a front (FL), oblique (OL) and side-view (SL) while the performance of walking along curved trajectories was evaluated using gait images captured by the front-view camera (FL). Walking along the big circle is labelled as Circle1 in the experiments while walking along the small circle is labelled as Circle2. Fig. 6.5 shows several silhouettes in one gait cycle of a subject walking along a curved trajectory in both gait datasets.



(a) Large Soton 3D-2D dataset

(b) KY4D dataset [44]

Figure 6.5: Curved walking.

6.3 Recognition results

In the section, we analysed the performance of our technique for the two types of walking (straight and curved) using two gait datasets: large Soton 3D-2D dataset and the KY4D dataset

[44]. We used the same strategy as explained in section 5.5 to evaluate the recognition results.

6.3.1 The results on the large Soton 3D-2D dataset

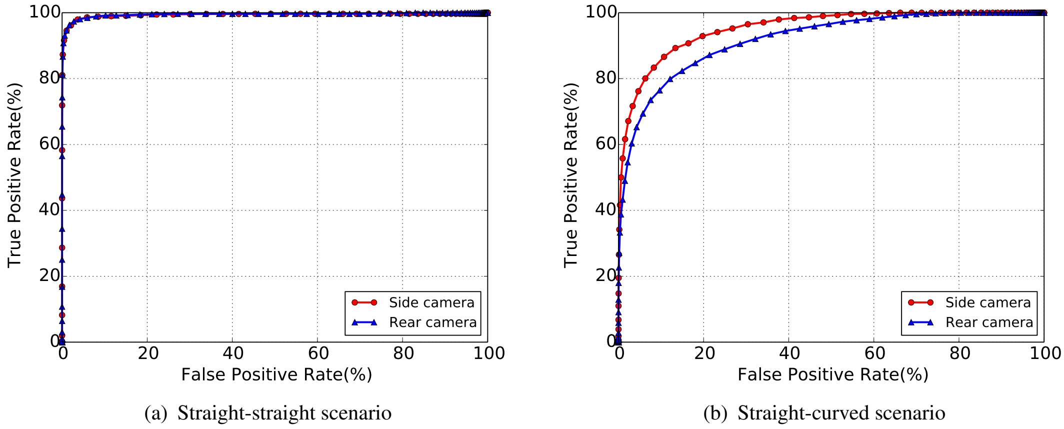
We investigated two scenarios for the gallery and probe in our dataset: straight-straight (i.e. straight walk data in the gallery and straight walk data in the probe) and straight-curved (i.e. straight walk data in the gallery and curved walk data in the probe). In each scenario, the performance was measured with respect to the number of divided areas in a silhouette and the number of frames per gait cycle.

Firstly, to measure the recognition rate with respect to the number of divided areas in a silhouette, we set the number of frames in a gait cycle to 30 because this number matches the average gait cycle length captured at 30 frames/second. Table 6.1 shows the Correct Classification Rate (CCR) for the two matching scenarios. The performance of the straight-straight scenario is generally high for different numbers of divisions and the best result is obtained at 4 divisions. As the number of divided areas in a silhouette increases, the recognition performance improves slightly as detailed information from different body parts improves the discriminatory power of the GFDs and ultimately enhance the recognition performance. However, when the number of divisions is greater than 4, the recognition performance decreases because much more detailed features are affected by noise which distorts the overall recognition performance. The number of divisions has more influence on the rear camera than on the side camera since the recognition rate decreases below 90% when the number of divisions is greater than 4. The performance of the straight-curved scenario is lower than that of the straight-straight scenario but the results take the same trend and the best performance is still observed at 4 divisions.

Table 6.1: CCR(%) versus number of divided areas in a silhouette in the large Soton 3D-2D dataset.

Straight-straight scenario					
Camera	**Number of divided areas**				
	1	**2**	**4**	**6**	**8**
Side	95.6	96.3	**99.6**	98.3	97.6
Rear	93.3	94.6	**98.3**	82.3	75.0
Straight-curved scenario					
Camera	**Number of divided areas**				
	1	**2**	**4**	**6**	**8**
Side	80.3	82.6	**86.0**	84.0	65.6
Rear	73.0	74.6	**77.3**	69.3	55.3

We calculated the 3-fold cross-validation for the best results (i.e. at 4 divisions). The average CCRs are 99.6% $\mp$ 0.5 (side camera) and 98.3% $\mp$ 1.7 (rear camera) for the straight-straight scenario while they are 80.6% $\mp$ 2.1 (side camera) and 73% $\mp$ 4.3 (rear camera) for the straight-curved scenario. We also draw the ROC curves in Fig. 6.6 for the 4 divisions. The verification rates of the first matching scenario are also better than those of the second scenario for both cameras. Furthermore, we measured the contribution of the upper half and lower half of the body for the straight walking where the CCRs are 80.3% and 97.3% respectively for the side camera and are 41.6% and 96.6% for the rear camera. These results reveal that the lower body region (leg motion) contributed more than the upper region towards the final recognition rate.



(a) Straight-straight scenario

(b) Straight-curved scenario

Figure 6.6: The ROC curves for the best matching results on the large Soton dataset.

Secondly, we measured the performance when the number of divisions is 4 and the number of frames in a gait cycle is varied from 5 to 30. We noticed that there is no improvement in performance beyond 30 frames. This is probably because our sampling procedure adds little additional information when the desired length is greater than the actual length. Therefore, in Table 6.2 we displayed the recognition results of the straight-straight and straight-curved scenarios versus the number of frames up to 30. From this Table, we can notice that the performance is not significantly influenced by the number of frames in a gait cycle for both matching scenarios. In the straight-straight scenario, the recognition performance is high for different numbers of frames and the maximum degradation in performance is around 3% when only 5 frames are used per gait cycle. However, the performance of the straight-curved scenario is lower than that of the straight-straight scenario and the maximum drop in performance is around 10% when only 5 frames are used per gait cycle. The estimated error in the two experiments when the number of correctly classified samples changes by one is about 0.3 percentage points.

6.3.2 The results on the KY4D gait dataset

In this section, we study the recognition performance as a function of the number of subdivisions of the silhouette and the number of sampled frames per gait cycle in the KY4D gait dataset. In the first experiment, we measured the variation of performance against the number of divided areas in a silhouette. Here, we fixed the number of frames at 20 in each gait cycle as this number matches the averaged length of gait cycle captured at 20 frames/second. Table 6.3 summarises the recognition results for both walking scenarios. We notice that the recognition performance

Table 6.2: CCR(%) versus number of frames/gait cycle in the large Soton 3D-2D dataset.

Straight-straight scenario						
Camera	**Number of frames**					
	5	**10**	**15**	**20**	**25**	**30**
Side	98.3	99.3	99.3	99.3	99.6	99.6
Rear	95.6	96.0	98.0	98.3	98.3	98.3
Straight-curved scenario						
Camera	**Number of frames**					
	5	**10**	**15**	**20**	**25**	**30**
Side	75.6	84.0	84.0	85.0	85.0	86.0
Rear	72.0	75.6	76.3	76.3	77.0	77.3

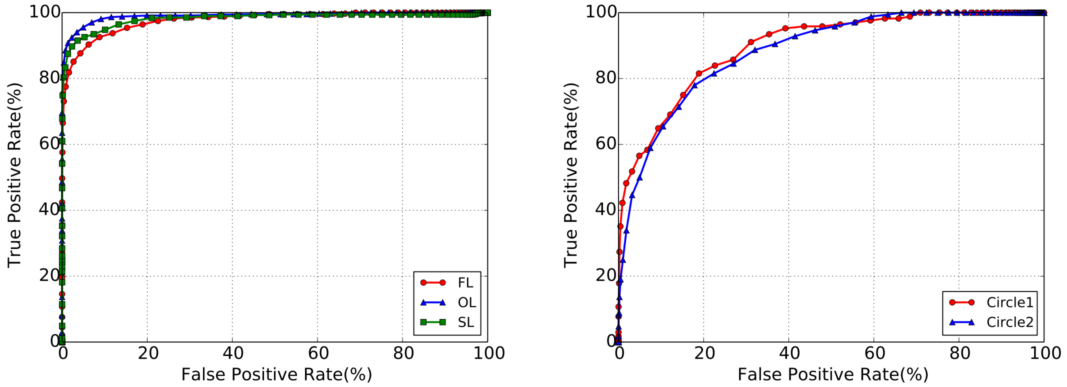
varies according to the camera viewpoint and the number of divided areas in a silhouette. For the straight-straight scenario, The number of divisions significantly affects the performance for the side-view camera while it has a lower influence for the oblique- and front-view cameras. The recognition rate is above 90% for the latter cameras using different numbers of divisions while it drops below 90% when the number of divisions is one and eight for the side-view camera. The reason for this may be because some people change the way they swing their arms and this is more evident from the side view than from other views. Generally, the best performance is observed at 6 divisions.

In the straight-curved scenario, the performance of gait recognition for walking along curved (circular) trajectories is lower than that for straight lines. The recognition rate is influenced by the number of divisions in a silhouette for both circular trajectories. The best performance for walking along a big circle (Circle1) is achieved at 6 divisions while for walking along a small circle (Circle2) is obtained at 1 division. The recognition rate deteriorates significantly when a higher number of divisions (i.e. 8) is used for the Circle1 and Circle2. Furthermore, walking along a small circle shows a lower performance than walking along a large circle. This might be due to the extreme curvature of the small circle, which could affect gait patterns of walking subjects accordingly. The average CCR for the best results is 67.8% while the difference in performance is $\sim$ 16% for the two circular trajectories. To illustrate the effect of walking along circular trajectories on the ROC curve and the class distribution, we plotted them for the two scenarios using the best matching results. Fig. 6.7 shows how the area under the ROC curve gets smaller for walking along circular trajectories. The class distribution has been plotted for one case in each scenario in Fig. 6.8. As can be seen, the overlap between the intra- and inter-class distributions gets wider for the second scenario, which indicates that the ability of the algorithm to discriminate between subjects gets lower for an extreme curvature trajectory.

In the second experiment, we evaluated the recognition performance by varying the number of frames per gait cycle from 5 to 20. The number of divided areas in a silhouette was fixed at 6 for walking along straight lines and the big circle and at 1 for walking along the small Circle. Table 6.4 shows the relationship between the CCR and the number of frames per gait cycle for the two scenarios. In general, the results illustrate that the performance degrades when the number

Table 6.3: CCR(%) versus a number of areas in a silhouette in the KY4D dataset.

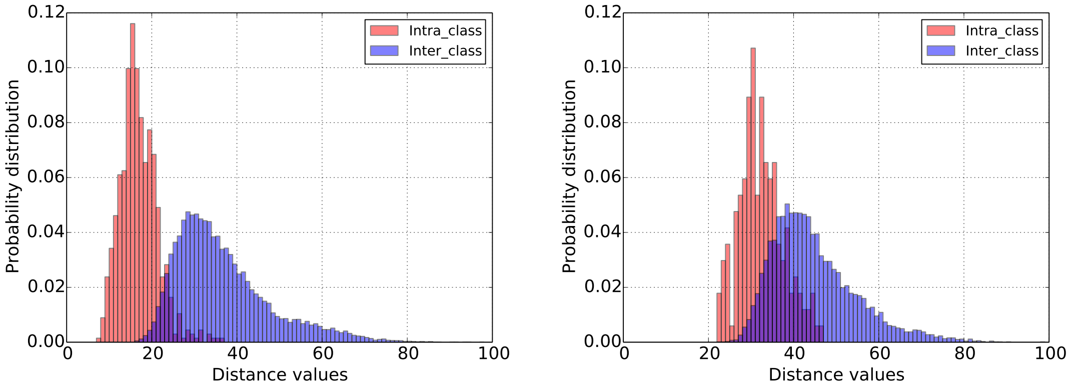
Straight-straight scenario					
Camera	**Number of divided areas**				
	1	**2**	**4**	**6**	**8**
FL	95.8	94.6	92.8	**97.6**	91.6
OL	95.8	97.6	97.6	**99.4**	98.8
SL	73.8	**97.6**	97.0	97.0	78.5
Straight-curved scenario					
Trajectory	**Number of divided areas**				
	1	**2**	**4**	**6**	**8**
Circle1	59.5	64.2	57.1	**76.1**	30.9
Circle2	**59.5**	57.1	40.4	54.8	33.3



(a) Straight-straight scenario

(b) Straight-curved scenario

Figure 6.7: ROC curves of the best results for the two matching scenarios on the KY4D dataset.



(a) Walking along a straight line (FL)

(b) Walking along a small circle

Figure 6.8: Class distribution for the two matching scenarios on the KY4D dataset.

of frames in a gait cycle decreases because the information available for recognition decreases. The degradation in performance depends on the viewpoint of the camera and the shape of the walking trajectory. For the straight-straight scenario, reducing the number of frames in a gait cycle has more influence on the SL camera than on the FL and OL cameras as the CCR goes below 90% when the number of frames reduces to 5 for the SL camera while it remains above 90% for the other two cameras. For the straight-curved scenario, The CCR significantly decreases from 76.1% to 57.1% when the number of frames is less than 20 for walking along the big circle (Circle1), which emphasises the importance of using sufficient number of frames for recognition, while the CCR gradually decreases from 59.5% to 54.7% when using 10 frames and then to 45.2% when using 5 frames for the Circle2. The estimated errors in the two experiments when the number of misclassified samples changes by one are 0.6% and 2.4% for the straight-straight and straight-curved scenarios respectively.

Table 6.4: CCR(%) versus number of frames per gait cycle in the KY4D dataset.

Straight-straight scenario				
Camera	**Number of frames/gait cycle**			
	5	**10**	**15**	**20**
FL	91.6	94.0	91.0	97.6
OL	94.0	97.6	98.8	99.4
SL	87.5	95.8	96.4	97.0
Straight-curved scenario				
Trajectory	**Number of frames/gait cycle**			
	5	**10**	**15**	**20**
Circle1	57.1	57.1	57.1	76.1
Circle2	45.2	54.7	59.5	59.5

The performance of the matching algorithm is also compared with the performance of the most related work in [44]. The comparison results are shown in Table 6.5 for the straight-straight and straight-curved scenarios respectively. The results demonstrate that the performance of the proposed algorithm in this chapter is favourably comparable to that in [44]. However, the algorithm depends only on three key frames in each gait cycle to do the 3D alignment and does not require a recursive implementation as does in [44].

Table 6.5: The comparison results for the straight-straight scenario in the KY4D dataset in terms of the CCR(%).

Approaches	**Straight-straight scenario**			**Straight-curved scenario**	
	FL	**OL**	**SL**	**Circle1**	**Circle2**
Proposed algorithm	97.6	99.4	97.6	76.1	59.5
Method [44]	99.4	96.4	98.2	71.4	61.9

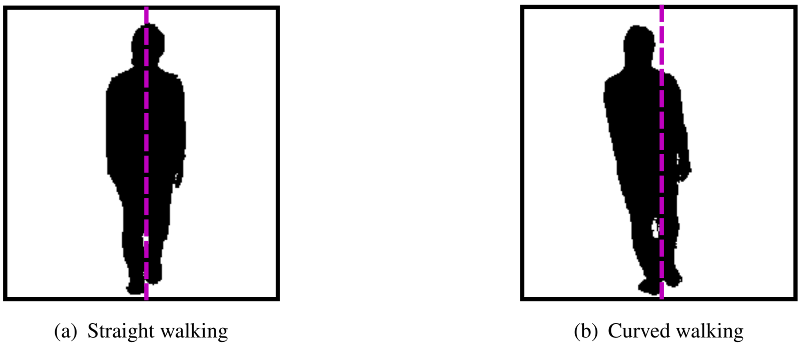
6.3.3 Matching curved with curved walking

The analyses using the large Soton and KY4D datasets in section 6.3.1 and 6.3.2 respectively showed that recognition results for the straight-straight scenario are much better than those for the straight-curved scenario where the difference in the CCR between the two matching scenarios was about 17% and 30% on the former and latter datasets respectively. We visually inspected the two types of walking and noticed some differences which have also been confirmed in several studies [108, 19, 18]. In curved walking, (1) the body leans towards the centre of the curvature, (2) the feet move asymmetrically where the inner foot twist more than the outer foot and (3) the head rotates into the direction of the future turn before the trunk. Some of these differences are illustrated in Figs. 6.9 and 6.10. These differences could modify the manner of walking for people according to the performed trajectory and imply that a direct comparison of the straight walk with the curved walk data may not be appropriate. We believed that using the curved walk data in the gallery may improve the performance with curved probes. Therefore, we measured the performance when matching curved galleries with curved probes. For the large Soton 3D-2D dataset, we set the number of frames in a gait cycle to 30 and the number of divided areas in a silhouette to 4. The CCR for the side and rear cameras using the whole dataset were 99.6% and 98.3% respectively.

The performance for the KY4D dataset was not recorded by [44] for this type of matching. To evaluate the performance for this dataset, we used 20 frames per gait cycle and 6 divisions in a silhouette. Since there is only one sequence for each of the two circular paths, we used the sequences for the small circle to identify the big circle sequences while the sequences for the big circle were employed to identify the small circle sequences. The CCR for the front-view camera (FL) were 80.9% and 76.1% for the first and second cases respectively. The recognition performance of this type of matching is better than that of straight with curved matching. However, the improvement is not as great as that found in the large Soton dataset. The difference in performance on the two datasets may be due to (1) a number of training sequences for each subject in the gallery where the KY4D includes only one sequence while the Soton dataset contains six, (2) different circular trajectories used in the gallery and probe in the KY4D as compared to the similar curved trajectories used in both gallery and probe in the large Soton dataset and (3) a lower frame rate of the KY4D in comparison with the large Soton dataset. Moreover, the estimated error in the CCR value when the number of misclassified sequences changes by one is higher in the KY4D than that in the large Soton dataset due to the smaller number of test sequences (i.e. 42 and 300) in the probe.

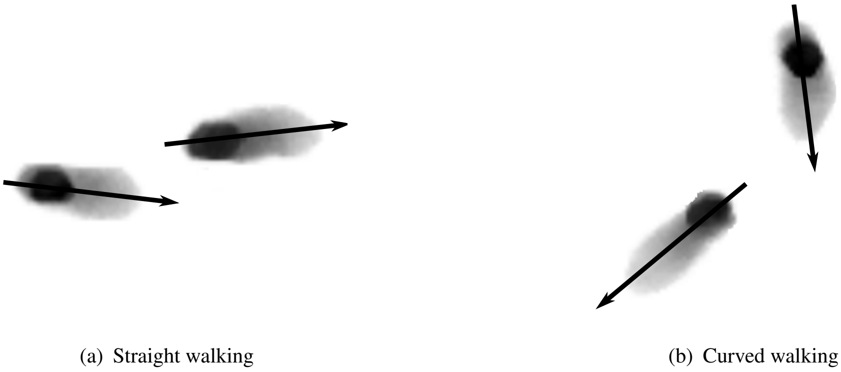
6.4 Discussion

In this chapter, a new matching algorithm has been proposed to recognise people regardless of their walking direction or viewpoints of cameras. The algorithm depends on three key frames at the start, middle and end of each gait cycle and assumes the motion to be linear between these key frames. The results show that matching straight with straight walking has a high performance while matching straight with curved walking achieves a lower performance on the large Soton 3D-2D and KY4D gait datasets. The algorithm also performed worse as the curvature of the walking trajectory was increased. The reason for the difference in recognition performance for


(a) Straight walking

(b) Curved walking

Figure 6.9: How the body pose differs during a curved walking, when viewed from a straight ahead (from the KY4D dataset).


(a) Straight walking

(b) Curved walking

Figure 6.10: The rotation of the two feet (from the KY4D dataset).

the two matching scenarios may be caused by a change in gait patterns of the subjects when walking along different trajectories, as shown in Figs. 6.9 and 6.10. The leaning of the body, the movement of the two feet and the rotation of the head and trunk vary according to the shape of the walking trajectory.

Further results illustrate that the performance of matching curved with curved walking is as high as that of straight with straight walking on the large Soton 3D-2D gait dataset but it is only a little better than that of straight with curved walking on the KY4D gait dataset. Moreover, the variation in performance with the number of frames in a gait cycle and the number of divisions in a silhouette is different for the two gait datasets. We believed that the differences are due to a variation in the characteristics of the two datasets arising from the different experimental setups. The two datasets were recorded using different cameras, frame rates and recording sites (i.e. narrow corridor and wide circular studio). The number of walking sequences for each subject and the shape of the walking trajectory are also factors. In the KY4D dataset, people have a smaller number of walking sequences and walked along constant curvature trajectories (i.e. circles) as compared to that in the large Soton dataset. This implies that the matching algorithm is data-dependent and its parameters need to be tuned accordingly. The results demonstrated that the proposed algorithm is compared favourably with the work in [44]. However, our algorithm does not require an estimate of the foot position from each frame or perform a recursive algorithm to precisely align each frame in a gait cycle.

We hypothesise that the discrepancies in gait patterns due to changing walking trajectories could affect the global shape of walking subjects. Therefore, we decide to investigate the performance of another gait features that depend on the motion as well as the shape in the next chapter to see what is the influence of this type of features on matching straight with curved walking.

Chapter 7

Are Simple Width-based Features a Solution to the Curved Walking Problem?

The performance of matching straight with curved walking in the previous chapter was low. The low performance was due to the discrepancy in gait patterns of people walking along different trajectories. This could affect the discriminatory power of gait features derived only from shape information (i.e. GFDs). We aim in this chapter to investigate the performance of other gait features which are sensitive to the motion of the body. The width of the outer boundary of the silhouettes [52] can capture the dynamic motion of the body as the temporal movements of different body parts can be represented. This feature also captures the physical structure of the body and represents the gait of a person in a compact way. However, this feature is not invariant against geometrical transformations such as rotation, translation and scale. Therefore, in order to accurately calculate width features, the synthetic silhouettes generated from 3D volumes of people should be precisely aligned with the real silhouettes. This requires that foot positions should be estimated from each frame in a gait cycle. The previous foot estimation procedure from 2D images in chapter 6 is applied only on the three key frames, when the two feet are close to each other and are approximately on the floor. The position is estimated by projecting foot points onto the floor when the position of the floor is known. However, when one of the two feet is in the air, there is an error in foot position estimation. In this chapter, we are not interesting in calculating foot positions. Therefore, we decided to calculate foot positions of subjects from each 2D image in one gait cycle from their 3D volumes. After aligning 3D volumes and generating synthetic silhouettes, width features can then be calculated from each synthetic and real silhouette along one gait cycle. The details of calculating width features and the analysis of their discriminatory power are given in this chapter.

7.1 3D Alignment process

The aim of the 3D alignment process is to remove the variation in appearance between the synthetic silhouettes generated from 3D volumes and real silhouettes from the camera. In this chapter, the 3D alignment is computed based on foot positions estimated from each frame along one gait cycle in a pair of a gallery and probe. The foot positions of 3D volumes in the gallery gait cycle are calculated using the same procedure described in section 6.1.2.2. The foot positions of the subject from 2D images in the probe gait cycle are estimated from their reconstructed 3D volumes in a similar manner. To remove the effect of outliers, we fit a second order polynomial to the estimated foot positions in both 2D and 3D gait cycles. After that, the local walking direction at each frame in both 2D and 3D gait cycles is calculated from differencing the fitted function. The local walking direction at each frame in a 3D gait cycle is calculated as follows: Assume F_i and F_{i-1} are the foot positions at a frame i and $i-1$, the local walking direction (δ_i) at a frame i is derived from the vector $\overrightarrow{D_i} = F_i - F_{i-1}$ such that

$$\delta_i = arctan\frac{D_{i_x}}{D_{i_z}} \tag{7.1}$$

The local walking direction at each frame in a 2D gait cycle is computed in a similar way. To align volumes in the 3D gait cycle according to the positions and walking directions of the subject in the 2D gait cycle, we normalise the number of frames in all 3D and 2D gait cycles as described in section 6.1.3 so that their phases are matched. We then set up a 3D coordinate system at each foot position along one 3D gait cycle and rotate each volume around its vertical axis by $\Delta\phi$

$$\Delta\phi = \varphi_i - \delta_i, \quad i = 0, 1...N-1 \tag{7.2}$$

where φ_i is the walking direction of a 2D silhouette at a frame i and N is the normalised number of frames in a gait cycle. Next, each rotated volume is translated by ΔF

$$\Delta F = f_i - F_i, \quad i = 0, 1...N-1 \tag{7.3}$$

where f_i is a foot position of a 2D silhouette at a frame i. Finally, each aligned volume is projected onto the 2D image plane of the probe camera to produce a synthetic silhouette with a similar appearance to that of a real silhouette from the camera.

7.1.1 Gait features calculation

In this chapter, we explore the discriminatory power of the width of the outer boundary of the silhouette [52] as a gait feature. This feature can capture the static (body height and width) and dynamic aspects (upper and lower limbs swing) of a human's gait. By extracting the width features from each silhouette along one gait cycle, the temporal movement of different body parts can be represented. These features showed a high performance for gait recognition [52, 23, 53, 93]. In a similar manner, Liu et al. [70] computed frieze features by counting the number of foreground pixels at each row and column of the silhouette over time. A pair of 2D arrays were calculated as gait features. After that, Tan et al. [111] expanded the frieze features by computing the normalised version over the two main diagonal as well as the horizontal and vertical directions. The normalised frieze features were calculated after resizing all the silhouettes to the same height. Then, the resulting features from each direction were averaged over the number of frames in a gait sequence as a final template for recognition. The performance of the normalised frieze features was evaluated individually from each direction. Inspired by the width features and the computation of the frieze features along four directions, we propose new gait features by computing the width of the outer contour of the silhouette along the four main directions: → horizontal (H), ↓ vertical (V), ↘ left diagonal (LD) and ↙ right diagonal (RD). These features can detect the movement of an individual from different directions along a gait cycle. After that, the performance is evaluated either from the individual features separately or by fusing features from multiple directions.

In order to compute the directional width features, silhouette images in one gait cycle are centred and cropped by placing a smallest square bounding box that covers the largest synthetic silhouette in the gallery to impart translation invariance. The width features along the required direction can then be extracted from each silhouette. Thus, four types of width features (called multi-directional width features) are calculated from each gait cycle. For a horizontal direction, the width feature is defined as a distance between the position of the first and last foreground pixels in each row. The remaining width features are computed in a similar manner by rotating the silhouette into the required directions. We noticed from the recognition results in chapters 5 and 6 that the performance is improved by fusing matching results of different features at the score level. Therefore, we decide to fuse the matching results of width features from multiple directions at the score level in a similar manner. However, the resulting features from the four directions will have different ranges depending on the physical structure of the body (i.e. person's width, height). This means that features with high values will dominate the effect of features with low values. To solve this problem, we normalise the feature vector from each direction by its maximum value. This will also make the extracted features scale invariant in their derived directions. Fig. 7.1 illustrates how the features are derived from the four directions along the silhouette image. The features before and after normalisation process are also presented where all the normalised features have the same range along the vertical axis.

Moreover, the orientation of the silhouettes in the large Soton dataset varies from frame to frame. In order to cope with small orientation changes due to the 3D alignment and projection processes, the feature vector extracted from one direction will be compared against the feature vector from the neighbouring (nearest) directions. For example, if we suppose that the horizontal direction represents $0°$, its neighbouring directions will be $\mp 2°$, $\mp 4°$, $\mp 6°$, $\mp 8°$. In the experiments, the vertical, left diagonal and right diagonal directions represent $90°$, $45°$ and $-45°$ respectively.

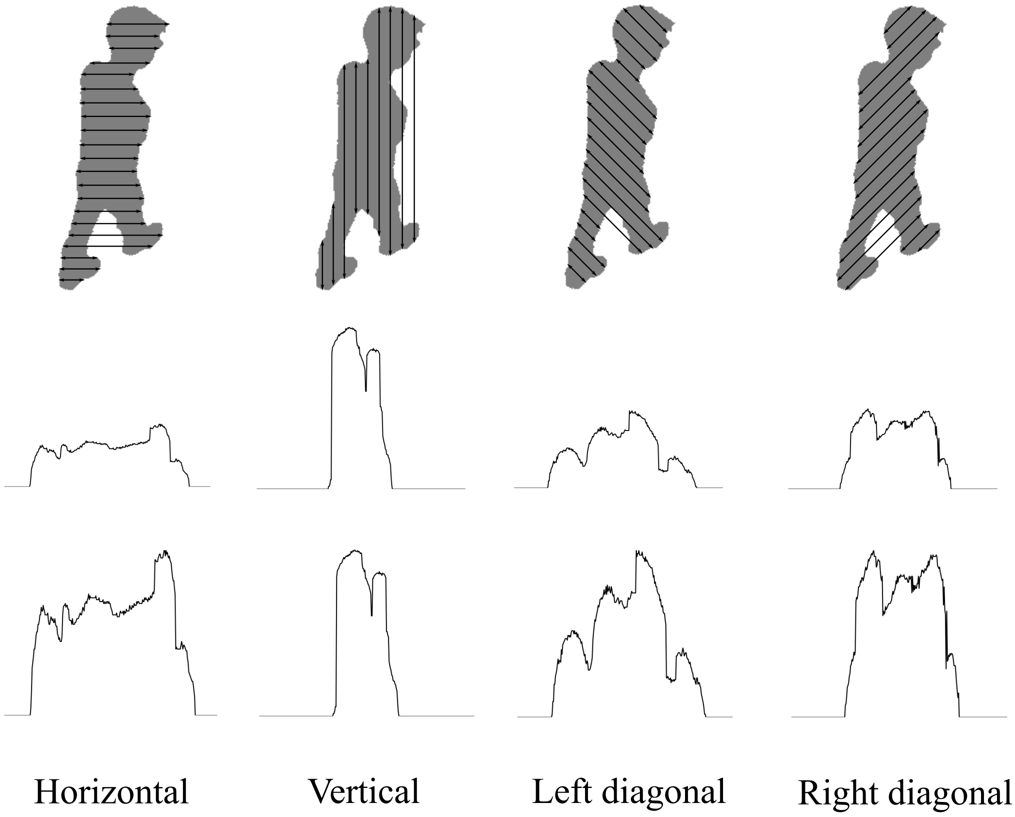


Figure 7.1: Sample of the multi-directional width features where the first row involves the directions along the silhouette, the second row contains the original features and the third row shows the normalised features.

7.1.2 Gait features matching

The following procedure is proposed to match gait features from a pair of a gallery and probe gait cycles. First, the features extracted from a frame i along a direction k in a probe gait cycle will be compared against the features extracted along the same direction from the neighbouring frames within a sliding window W and also from the nearest directions N in each of these frames in a gallery gait cycle. Secondly, the minimum distance is selected for the frame i along the direction k as

$$M_{i,k} = \min(||p_{i,k} - g_{i+j,k+\Delta r}||),\ j \in W,\ \Delta r \in N \tag{7.4}$$

where $||.||$ is the Euclidean distance, p and g are the probe and gallery gait feature vectors respectively, i and $i+j$ are the frames indices, k is one of the main directions $\{0°, 90°, 45°, -45°\}$ A set of ranges in a window W is $\{0, \mp1, \mp2, \mp3\}$ while a set of neighbouring directions N is $\{0, \mp2°, \mp4°, \mp6°, \mp8°\}$. Thirdly, the total distance between two gait cycles of length L along the direction k is computed as

$$T_k = \sum_{i=0}^{L} M_{i,k} \tag{7.5}$$

Finally, the sum rule is used to fuse the results of matching gait features from multiple directions at the score level through

$$D = \sum_k T_k, \quad k \in \{0^\circ, 90^\circ, 45^\circ, -45^\circ\} \tag{7.6}$$

Using eq. 7.6, different combinations of features can be fused together as will be illustrated in the results section.

7.2 Recognition results

The effectiveness of the multi-directional width features for gait recognition on straight and curved trajectories is illustrated below. We used the same gait datasets as we used in chapter 6 (large Soton and KY4D). First, the effect of different parameters used in the features matching procedure is analysed. Then, the recognition results of the individual features and the fusing of different combinations of features are demonstrated. As explained in section 6.3, we set the number of frames in all gait cycles to 30 and 20 in the Soton and KY4D datasets respectively because these numbers are matched the average gait cycle length captured at a rate of 30 and 20 frames/second respectively. Two scenarios are investigated in the experiments: matching straight walking in the gallery with straight walking in the probe (straight-straight) and matching straight walking in the gallery with curved walking in the probe (straight-curved). The same probe cameras employed in chapter 6 to evaluate the performance of gait recognition on straight and curved trajectories in both datasets are used in this chapter. The positions of these cameras are shown in Figs. 6.3 and 6.4. The performance of the normalised width features is compared with the performance of the normalised frieze features over the four main directions. The frieze features over the horizontal direction is calculated as $C_H(y,t) = \sum\limits_x S(x,y,t)$ where S is a silhouette and (x,y,t) is a specific pixel location (x,y) at a frame t. The frieze features over the other directions are computed in a similar manner by rotating the silhouette into the required directions (90° for vertical direction, 45° for left direction and $-45°$ for right diagonal direction).

7.2.1 The effect of the range of the window for matching neighbouring frames

In this experiment, we analysed the performance versus different ranges of the window (W) used in Eq. 7.4. Fig. 7.2 shows the recognition results when fusing all the width features for matching straight with curved walking on both datasets. In the large Soton dataset, gait images of people walking along a curved trajectory were captured by the rear camera. The results show that increasing the size of the window W when matching a frame in the probe with its neighbouring frames in the gallery has a small influence on the performance where the CCR slightly increases as the size of the W gets wider. For the KY4D dataset, gait image were captured by the front-view camera (FL) for the people walking along the big circle. The results show a significant impact of the parameter W on the performance where the CCR rises from 66.6% to 78.5% at $W = \mp1$ (i.e. 3 frames comparisons) and to 80.9% at $W = \mp2$. However, no further improvement is found when $W > \mp2$.

The reason for the variation in the effect of the parameter W on the two datasets might be due to a variation in frame rates and curvature of the walking trajectories. The frame rate is smaller and people walk along a more extreme curvature trajectory in the KY4D dataset than in the large Soton dataset. This could increase the influence of an error in the gait cycle or walking direction estimation on the recognition performance when matching corresponding frames directly. The size of the test samples in the dataset has also an impact on the performance as the variation in the CCR when the total number of the correctly classified samples changes by one is about 0.3% for the large Soton dataset and around 2.4% for the KY4D dataset. Based on these results, we set W at ∓2 in the remaining analyses because increasing the window size beyond ∓2 for the large Soton dataset does not lead to any significant gain.

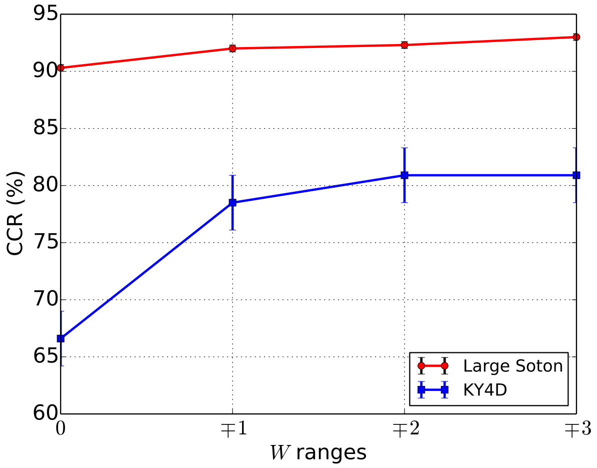


Figure 7.2: CCRs of fusing all width features for different ranges of W when matching straight with curved walking.

7.2.2 The influence of matching nearest directional features

We conducted this experiment to measure the influence of matching gait features derived from a specific direction in the probe with the features from its nearest directions in the gallery over a range of intervals (N) from $\mp2°$ to $\mp8°$. The same probe camera and test curved trajectory of the previous experiment were used here for both datasets. The results of matching straight with curved walking are indicated in Fig. 7.3 when fusing all the width features. For the large Soton dataset, the performance fluctuates upward as the interval of matching nearest width directional features expands. A possible explanation for the fluctuation is that the best match at a larger interval of nearest directions should be at least as good as or better than the match at a smaller interval but the best match could be made for the incorrect person in the gallery according to their distance. In the KY4D dataset, the effect of the parameter N is limited to a smaller interval (i.e.

$\mp 2^\circ$). The CCR increases from 80.9% to 83.3% at $N = \mp 2^\circ$ and then returns to 80.9% as N takes higher intervals.

These results reveal that the parameter N has a more influence on the large Soton dataset than on the KY4D dataset. This might be attributed to orientation changes of silhouettes in the former dataset which could be better captured using a larger interval of nearest directional width features. The initial analyses were done only for N = 0°, $\mp 4^\circ$ and $\mp 8^\circ$ due to the long processing time requirements. According to the corresponding results at that point, we set N at $\mp 8^\circ$ in the remaining analyses in this chapter because this would provide a balance between a good performance and a reasonable computational time. Subsequently, the results of the intermediate intervals reveal that the performance slightly varies up and down on the large Soton dataset. However, we do not believe that there is a significant improvement in the CCR for a larger interval of neighbouring directional features. For the KY4D dataset, the bump in Fig. 7.3 is well within the estimated error and has been ignored.

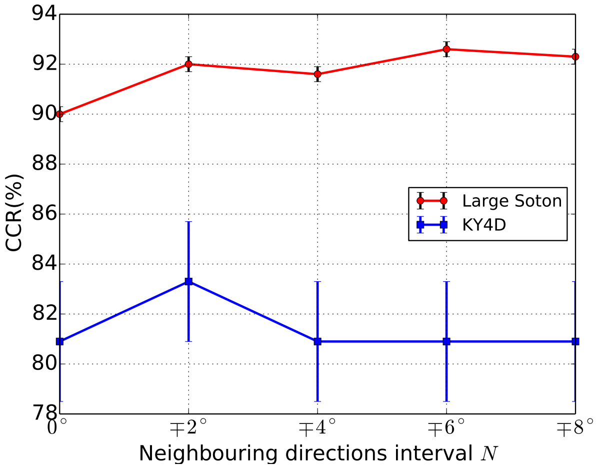


Figure 7.3: CCRs obtained by fusing all width features for different intervals of neighbouring directional features N when matching straight with curved walking.

7.2.3 The influence of fusing features

The aim of this experiment is to evaluate the performance of the individual width features and the fusing of different directional features using Eq. 7.6 for both walking trajectories (i.e. straight and curved). All gait features are matched using 2 frames comparison (i.e. each frame in the probe gait cycle is compared with the corresponding frame, two frames before and two frames after in the gallery gait cycle). We also evaluated the performance of the normalised frieze features on both datasets. To fairly compare the performance of the normalised frieze with the normalised width features, we used Eq. 7.6 to fuse the features from multiple directions. Table 7.1 shows the performance on the large Soton 3D-2D dataset using different gait features. As can be seen, the performance of matching straight with straight walking is still higher than that of matching straight with curved walking. The possible reason for this, as explained before, is the discrepancy in gait patterns of subjects walking along different trajectories. However, we notice a high performance for the latter scenario from the rear camera using the width and frieze features where the recorded CCR is above 94%.

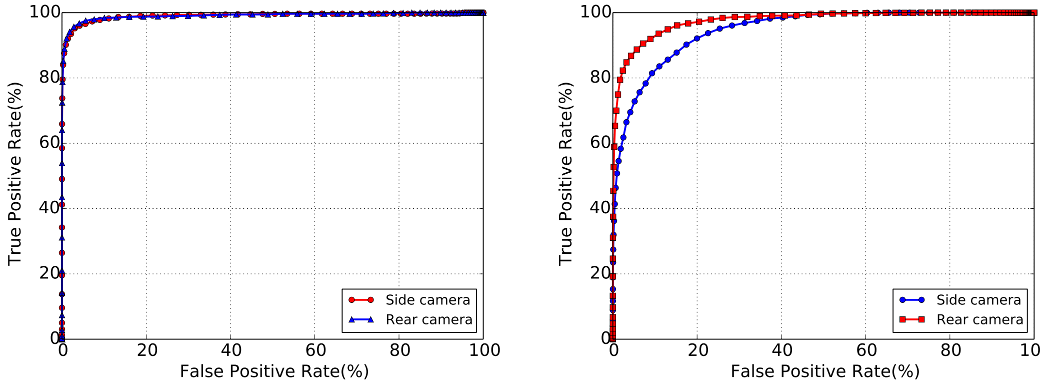
In general, the recognition results of the individual directional (width and frieze) features reveal that the horizontal (H) and left diagonal (LD) features perform better than the vertical (V) and right diagonal (RD) features. The possible explanation for this behaviour is that the movements of different body parts could be best represented using these features from the selected viewpoints. Fig. 7.1 shows how the horizontal and left diagonal features capture more information that describes different body parts of a subject than their counterparts from the vertical and right diagonal direction. Further results demonstrate that the fusion strategy improves the recognition performance of the individual features since more evidence is available for recognition. The influence of the fusion strategy is more prominent when matching straight with curved walking. The change in walking direction at each frame could distort the discriminatory power of the features extracted from a specific direction in the 2D image as the temporal movements of body parts could not effectively be captured in this scenario. By combining features from multiple directions, more discriminant information could be provided, which eventually improves the overall recognition performance. We also notice that the performance of the width features is similar to that of the frieze features since both of these features capture the same type of static (physical body structure) and dynamic (motion dynamics of the body) information. Furthermore, there is a variation in the performance for fusing different sets of features according to the viewpoint and performed trajectory.

Due to the performance similarity of the two directional features (i.e. width and frieze), Fig. 7.4 shows only the ROC curves obtained by fusing all the width features for the two matching scenarios. As can be seen, the verification rates of matching straight with straight walking are higher than those of matching straight with curved walking. The average CCRs of the 3-fold cross-validation for fusing all the width features are $98.6\% \mp 0.9$ (side camera) and $99.0\% \mp 1.4$ (rear camera) for the straight-straight scenario while they are $76.3\% \mp 3.1$ (side camera) and $92.0\% \mp 3.7$ (rear camera) for the straight-curved scenario.

To show which process (i.e. 3D alignment, new directional features) contributes more to the recognition performance improvement, we report in Table 7.1 the performance of the GFDs with the new 3D alignment (GFDs1) and the best performance (GFDs2) from the previous 3D alignment in chapter 6. We evaluated the performance of the GFDs using 4 silhouette divisions and then calculating 60 features from each one, as explained in section 6.1.4. The features from each division are matched using 2 frames comparison and fused at the score level using the sum rule. The results demonstrate that the new 3D alignment only improves the matching of straight with curved walking from the side camera by about 4%. The performance of the width and frieze features from the side camera is similar to that of the GFDs for the straight-straight scenario but it is lower for the straight-curved scenario. However, for the rear camera the performance of the multi-directional features is better than that of the GFDs for both matching scenarios where the

Table 7.1: CCR(%) on the large Soton 3D-2D dataset.

Directional features	**Straight-straight scenario**				**straight-curved scenario**			
	Side		**Rear**		**Side**		**Rear**	
	Width	**Frieze**	**Width**	**Frieze**	**Width**	**Frieze**	**Width**	**Frieze**
H	96.3	96.6	98.0	98.6	70.3	67.6	81.0	83.6
V	89.3	90.3	88.6	97.3	51.6	50.3	58.6	68.6
LD	98.6	98.0	98.0	96.3	77.0	77.3	89.6	**94.6**
RD	88.0	88.3	95.6	97.6	48.3	47.6	76.6	76.3
All	98.3	98.0	99.0	98.6	81.6	82.0	92.3	92.0
LD+RD+H	98.6	**99.3**	**99.3**	**99.3**	79.6	78.3	94.3	92.0
LD+RD+V	97.0	98.3	98.6	98.6	79.6	78.3	84.0	92.6
LD+H+V	99.3	98.6	98.3	**99.3**	**82.0**	**82.6**	93.3	91.3
RD+H+V	95.6	96.3	98.3	98.3	75.0	71.0	87.3	88.6
H+V	96.0	96.6	98.0	98.3	77.0	76.0	83.6	85.6
H+LD	**99.6**	98.6	98.6	99.0	81.3	81.3	**94.6**	93.3
H+RD	96.0	96.3	**99.3**	98.6	68.3	59.6	85.6	85.6
V+LD	97.3	97.0	98.0	98.6	81.0	78.3	85.3	88.3
V+RD	94.0	96.6	97.0	97.6	61.0	59.6	82.3	85.0
LD+RD	97.6	98.0	99.0	99.0	79.3	79.0	90.6	92.3
GFDs1	99.6		98.6		90.6		77.3	
GFDs2	99.6		98.3		86.0		77.3	



(a) Straight-straight scenario

(b) straight-curved scenario

Figure 7.4: The ROC curves obtained by fusing all the width features for the two cameras in the large Soton 3D-2D gait dataset.

best recognition rates using the former features are over 99% and 94% for the first and second matching scenarios respectively. The reason could be that the movements of different body parts are better seen from the rear camera due to its position and orientation. In contrast to the rear camera, the tilt of the side camera and the closer distance between the subject and the camera make the change in the pose of a subject during curved walking more extreme than during straight walking and as a result the movements of body parts could not be captured efficiently by the width and frieze features.

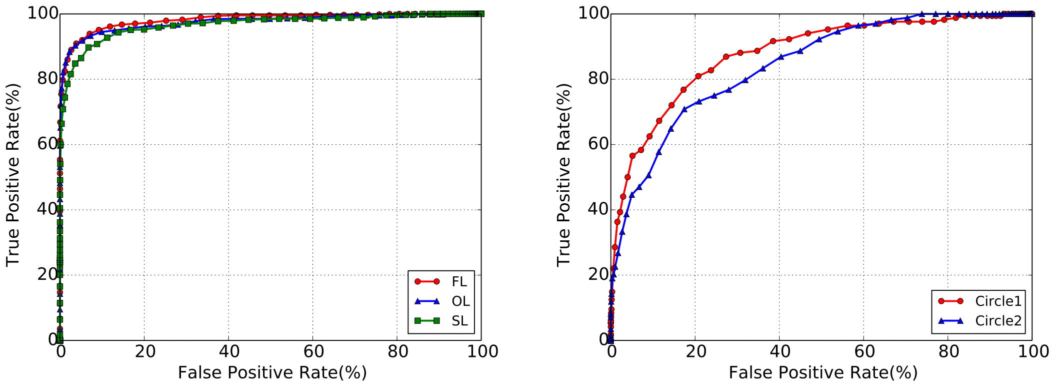
We also evaluated the performance of the width and frieze features on the KY4D gait dataset. For illustration, we only reported the performance of fusing all the features and the best performance obtained for both matching scenarios using the width and frieze features respectively in Table 7.2. As there is more than one combination of features that gives the best results, we draw only the ROC curves obtained by fusing all the width features for both matching scenarios in Fig. 7.5. The recognition results reveal that the performance of matching straight with straight walking is still better than matching straight with curved walking. We notice the following points for the best CCR values: first, the average difference is just around 19% for the CCR between the two matching scenarios. Secondly, the performance of the frieze features is better than that of the width features for straight walking and is similar for curved walking. This could be related to the movement of the limbs and its influence on the way the two features calculated. Thirdly, the difference in the CCR between the two features for walking along a straight line decreases (according to the viewpoint) from 1.8% to 0.6% as the camera moves from the side to the front view. Fourthly, the performance of walking along a small circle (Circle2) is lower by only about 10% than that of walking along a big circle (Circle1). This difference was probably because of the variation in the curvature of the walking trajectory, which may affect the discriminatory power of the features. Finally, we report in table 7.2 the performance of the GFDs with the new alignment and their best performance from the previous alignment in chapter 6. The number of divided areas in a silhouette was chosen based on the best results obtained in section 6.3.2. The results illustrate that the performance of matching straight with straight is similar to that of the previous alignment and matching straight with curved walking still performs worse.

Table 7.2: CCR(%) on the KY4D dataset for different walking trajectories where FL, OL, and SL refer to different views for the straight walking.

Features	**FL**	**OL**	**SL**	**Circle1**	**Circle2**
Width (all)	98.2	95.8	94.6	80.9	69.0
Frieze (all)	97.6	97.0	95.8	78.5	69.0
Width (best)	98.2	97.0	94.6	83.3	73.8
Frieze (best)	98.8	98.2	96.4	83.3	73.8
GFDs1	96.5	99.4	97.6	80.9	50.0
GFDs2	97.6	99.4	97.6	76.1	59.5

7.2.4 The effect of noise on the multi-directional features

The influence of using noisy silhouettes in the probe set on the performance of the multi-directional features is studied in this section. To do that, we added salt and pepper noise to the probe images from the oblique camera in the KY4D dataset. We then extracted the largest connected component object in each image. After that, the features were calculated from the synthetic and noisy images in the gallery and probe sets respectively. The noise rate varies from 10% to 40%. The corresponding CCRs obtained by fusing all the features are presented in Fig. 7.6. We can see that the performance of both width and frieze features is not influenced by lower



(a) Straight-straight scenario

(b) Straight-curved scenario

Figure 7.5: The ROC curves of fusing all the width features for both matching scenarios in the KY4D gait dataset.

noise levels (between 10% and 20%). However, when the noise level rises to 30%, the drop in the CCR is only 0.6% for the frieze features while it is around 12% for the width features. At higher noise rates (i.e. 40%), the performance of both features severely deteriorates. The experiment was done only once due to the longer processing time required. The estimated error in the CCR when the number of correctly matching samples changes by one is about 0.6%. These results reveal that the frieze features are more robust but there is a little difference for low noise levels.

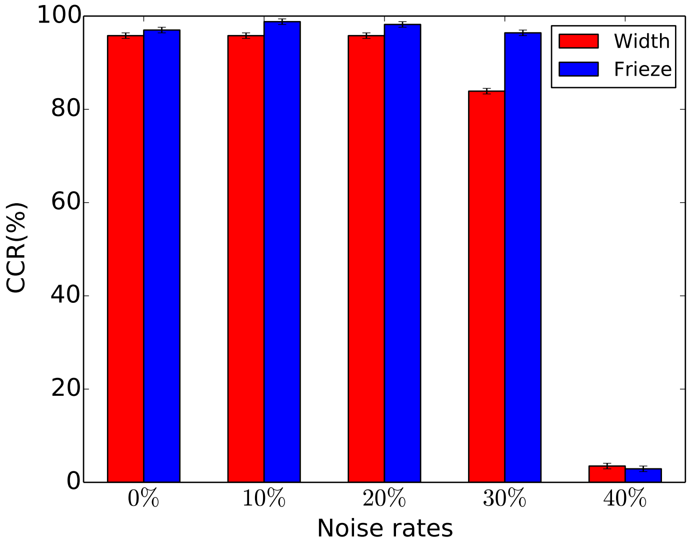


Figure 7.6: The effect of noise on multi-directional features performance.

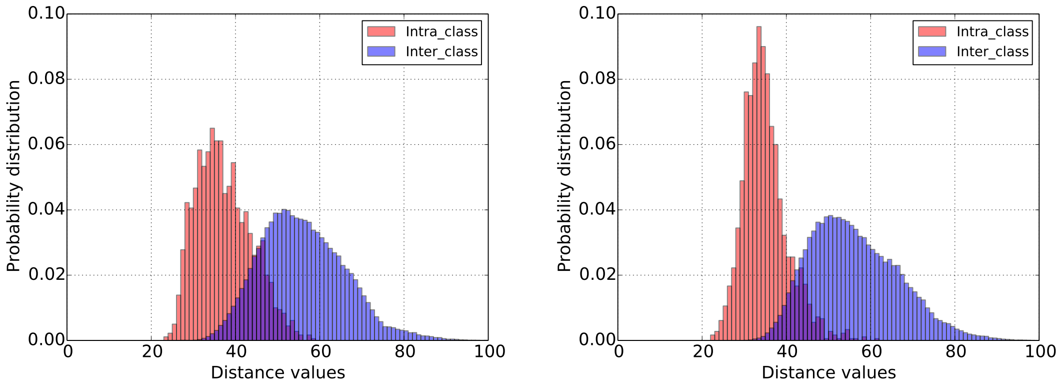
7.3 Discussion

This chapter showed the effectiveness of multi-directional width and frieze features for gait recognition on straight and curved trajectories. These features can capture the static (e.g. height and width) and dynamic (e.g. swings of body limbs) information of the walking subjects. The normalised version of these features is computed along four main directions in the image and matched separately by means of a matching strategy based on the neighbouring frames and nearest directions in each of these frames. Finally, gait features from multiple directions are fused at the score level by using the sum rule of their matching results. The performance of these features is evaluated on two gait datasets: large Soton 3D-2D and the KY4D.

The initial recognition results of the width features are shown to be better than the frieze features, which motived us to evaluate the performance of different combinations of these features from multiple directions. However, experiments with the complete set of fused combinations of different features revealed that there is no significant difference in performance between the width and frieze features. The influence of different parameters in the features matching procedure on performance varies on both datasets. This variation may be due to the different characteristics of these datasets. The final recognition results using the width and frieze features on both datasets indicate that the performance of matching straight with straight walking is still higher than that of matching straight with curved walking. However, an improvement in performance is obtained for the latter. We believed that the type of features and their fusing and matching strategy together have contributed to this improvement.

Fusing matching results of multiple directional features has a prominent role in improving the performance of matching straight with curved walking where the performance of the individual features for curved walking is inferior to that recorded for straight walking. A possible explanation is that the discriminatory power of the features extracted from one direction may be distorted when the walking direction changes at each frame as the movements of body limbs could not be efficiently captured by the unidirectional features. By combining features from multiple directions, more evidence is available for recognition.

The performance of matching straight with curved walking using the width and frieze features for the side camera is not as high as the rear camera in the large Soton 3D-2D dataset. A possible reason is the extreme change in the pose of the walking subject during curved walking, which makes it difficult to efficiently capture the temporal movements of different body parts. The camera is placed at the top of the capturing scene and there is a small distance between the camera and the subjects. Fig. 7.7 shows the class distribution obtained by fusing all the width features for matching straight with curved walking. We can see that the overlap between the intra- and inter-class distribution for the side camera is wider than that for the rear camera. We also notice that the performance in the KY4D dataset is affected by the extreme curvature of the walking trajectory, in that the performance dropped by 10% when the subjects walked along the small circle. The drop in performance may be caused by the discrepancy in gait patterns of subjects walking along different trajectories as explained in section 6.3.3. Moreover, no global set of multiple features



(a) Side camera

(b) Rear camera

Figure 7.7: The class distribution of fusing all the width features for matching straight with curved walking in the large Soton 3D-2D gait dataset.

gave the best performance for all test cameras and walking trajectories in the two datasets. This implies that the parameters of the algorithm should be tuned according to the viewpoint of the camera and the walking trajectory. Finally, both multi-directional width and frieze features are resilience to small amounts of noise. This could be because of the normalisation process in the feature extraction.

We can conclude from the results that the multi-directional width and frieze features do not solve the problem of matching straight with curved walking.

Chapter 8

Conclusions and Recommendations

8.1 Conclusions

Several gait-based human identification techniques have been proposed in the literature and have already achieved high recognition rates. However, most of these techniques implicitly assume that the observation angle within one gait cycle is approximately constant and that people walk only along straight lines. However, there are variations in the observation angles and orientations within one gait cycle even if people do not change their walking directions. This creates a change in a subject's pose and appearance from frame to frame. The effect of this change is prominent when the distance between the camera and the subject is small and the camera captures nearly side-view images. In reality, people also change their walking direction either from time to time or at each time instance in order to turn corners or avoid obstacles. In these situations, there are continual changes in the local walking direction (walk along curved trajectories). The change in subject's pose and appearance is more extreme during curved walking than during straight walking. This thesis has investigated these observation issues by enrolling people in 3D using a synchronous multi-camera setup. Matching was then performed against gait images from a single independent camera(s). Two different gait matching algorithms have been proposed for this purpose. Furthermore, the Generic Fourier Descriptors (GFDs), which are based on shape information, have been explored as gait features in the proposed matching algorithms. In the previous chapter, the performance obtained from using gait features that are sensitive to the dynamic motion of the body as well as the physical body structure (the normalised width and frieze patterns over four main directions) was evaluated for gait recognition on straight and curved trajectories. Finally, to support all the above, an expanded version of the Biometrics Tunnel has been developed to aid investigating the performance of matching 3D against 2D gait data obtained from independent (asynchronous) cameras.

The first set of experiments in chapter 3 investigated the possibility of matching 3D volumetric data against data from single cameras inside the Biometric Tunnel using the most popular gait representation, Gait Energy Image (GEI), as gait features. In these experiments, a subset of 43 subjects from the Soton multi-view gait dataset [98] was used and the influence of the viewing angle of the camera and the distance between the camera and the subject on recognition performance was evaluated. An average recognition rate of 97% was recorded from matching 3D volumetric data against gait images from single cameras at the far ends of the Tunnel where the observation angle and subject's orientation in one gait cycle was approximately constant while a low recognition rate of about 42% was obtained when the matching was performed against the middle (near) cameras. The analyses showed that the possible reasons for the modest performance of the latter were (1) the distortion in the 3D volumetric data when one camera was removed during the 3D reconstruction process and (2) the large variation in the observation angle and subject's pose in one gait cycle, which made the derived gait features (i.e. GEI) unstable. Based on the results of these analyses, the subsequent works were proposed.

The layout of the Biometrics Tunnel was modified and expanded, as explained in chapter 4, to give a system capable of capturing synchronous multi-view and asynchronous single-view gait images from two independent cameras. More specifically, the issues emerging in the old Tunnel system such as problems of synchronisation, lighting conditions and camera alignment, were resolved. An expanded configuration was devised with two additional wide-angle lens cameras, which were placed at different positions and heights to capture independent 2D gait images from a set of side and rear views. Using the new Tunnel system, two 3D-2D gait datasets were collected. The first small gait dataset includes 17 people (13 men and 4 women) walking along a straight line in the middle of the Tunnel. Each subject was recorded ten times by the cameras. Meanwhile, the second larger dataset comprises 50 people (33 men and 17 women) walking along straight and curved trajectories. Each subject walked six times along a straight line and six times along a curved trajectory in the middle of the Tunnel. The 3D volumetric reconstructions of people from the synchronous multi-cameras are built for the gallery set while the gait images from each of the two independent cameras are captured for the probe set.

In chapter 5, the first gait matching algorithm was proposed to identify people walking along a straight line with gait images captured by a perspective camera mounted on the wall in a narrow corridor (to simulate an access control scenario). The appearance of a person varies according to his/her position in the images. The proposed algorithm accommodated the problem of appearance variations by matching the position of each pair of 3D and back-projected 2D gait cycles in the gallery and probe sets respectively. The positions of the 3D and the back-projected 2D gait cycles in 3D space were estimated from the centroid of the first frame in each of these gait cycles such that the position of the back-projected 2D gait cycle was calculated by back-projecting the first silhouette onto the walking plane of the people in the gallery and calculating the centroid of the back-projected silhouette in 3D space while the position of the 3D gait cycle was computed from the centroid of the first 3D volume. After that, each 3D volume along one gait cycle was aligned by 3D translation along the walking direction and then projected onto a 2D image plane of the probe camera using its projection matrix to produce a synthetic silhouette. Next, gait features were extracted from each synthetic and real silhouette along one gait cycle using the GFDs and compared using Dynamic Time Warping. The performance of this algorithm was evaluated using the small Soton 3D-2D gait dataset collected in the Tunnel and a recognition rate of up to 98.8% was obtained. By dividing each silhouette into multiple areas and calculating gait features from each area for recognition, the discriminatory power of the GFDs from the side camera was significantly improved since different body parts can be well represented and more detailed features were provided.

Further investigations showed that samples were misclassified because of the people did not walk along the main straight line in the Tunnel and the GFDs had a low discriminatory power when the whole silhouette region was used for feature extraction. Several experiments were carried out to illustrate the effectiveness of the proposed algorithm using gait images captured by the rear camera. The results revealed that the GFD resolutions (i.e. the number of features) had a slight influence on recognition performance where the recognition rate was above 90% when a modest number of features per frame was used for recognition. The algorithm was also insensitive to salt and pepper noise and perspective distortion from different viewpoints on both sides of the Tunnel. The results also demonstrated that the algorithm can efficiently handle truncated gait cycles of different lengths in that the recognition rate did not drop below 95% even when only 25% of the total gait cycle length was available for recognition. However, calibration errors of the probe camera had a negative impact on performance. Moreover, the algorithm is not applicable when the optical axis of the camera is parallel to the walking direction (i.e. front/rear views) because back-projecting the silhouette onto the walking plane is ill-posed.

The main drawbacks of the previous technique are characterised and the appropriate solutions are devised by developing a second matching algorithm in chapter 6. The algorithm is (1) completely viewpoint independent and (2) not restricted to walking only along straight lines. In the second algorithm, matching each pair of 3D and 2D gait cycles was based on the foot positions at the start, middle and last frames in each of these cycles. By assuming the motion between these key frames to be linear, the intermediate foot positions and the local walking directions between the key frames are estimated. After that, each 3D volume along one gait cycle was aligned using 3D rotation and translation and then projected onto the 2D image plane of the probe camera to generate a synthetic silhouette. Next, subjects were identified using GFDs as gait features. To test the performance of gait recognition on straight and curved trajectories, the large Soton 3D-2D gait dataset and the Kyushu University 4D Gait Database (KY4D) were used. The results, in general, demonstrated that the performance of matching straight with straight walking was high where the average recognition rate of around 98% was achieved on both datasets. However, matching straight with curved walking was low where the average recognition rate for the best matching results were 81.6% on the large Soton 3D-2D gait dataset and 67.8% on the KY4D dataset. The discrepancy in gait patterns of people walking across different trajectories was the reason for the low performance. The leaning of the body, the twist of the two feet and the rotation of the head with respect to the trunk varied according to the shape of the trajectory. Further results showed that the performance of matching curved with curved walking was as high as that of matching straight with straight walking on the large Soton dataset while the performance of matching curved with curved walking was a little better than that of matching straight with curved walking on the KY4D dataset. The results also showed that the recognition performance improves as the number of divisions in a silhouette increases in the large Soton dataset but there is no improvement when the number of divisions is greater than 4. Furthermore, the performance improves as the number of frames in a gait cycle increases. The changes in recognition performance with respect to the number of divisions in a silhouette and the number of frames in a gait cycle do not take a similar trend in the KY4D dataset as those in the large Soton dataset. The reasons for the difference may be due to the variations in the characteristics of these datasets because their experimental setup was different. The two datasets were captured using different cameras, frame rates and recording sites. They also have a different number of walking sequences per subject and different walking trajectories. It can be concluded that the performance of the matching algorithm is data-dependent and its parameters need to be tuned according.

In chapter 7, we measured the influence of using gait features that are sensitive to the motion of the limbs as well as the physical body structure on matching straight with curved walking. The performance of the normalised width of the outer contour of the silhouette and the frieze pattern over four main directions (horizontal, vertical, left diagonal and right diagonal) was investigated. In order to compute these features, each volume in the gallery gait cycle was aligned according to the foot position and walking direction of the corresponding silhouette in the probe gait cycle and projected onto a 2D image of the probe camera to produce a synthetic silhouette with a similar appearance to the real silhouette. The features extracted from each frame over one direction were then compared against those derived from the nearest directions and from the neighbouring frames within a sliding window. A final matching result was computed by combining the matching results of the individual features from multiple directions at the score level using the sum rule. The recognition results indicated that the normalised width and frieze patterns have similar performance. The average Correct Classification Rates (CCRs) of the best matching results were 99.5% and 88.3% using the width features for matching straight with straight walking and matching straight with curved walking respectively on the large Soton dataset and were 96.6% and 78.6% on the KY4D dataset. The recognition results showed that matching straight with straight walking is still higher than matching straight with curved walking. However, an improvement in the performance had been obtained for the latter using the width features on both datasets. This means that aligning the volumetric data using 3D rotation and translation only may not be sufficient enough to simulate the orientations of the body parts during curved walking. Instead, a more sophisticated alignment procedure should be developed according to the curvature of the walking trajectory. This may include

1. how much each of the inner and outer foot should be twisted?,
2. how should the leaning of the body towards the inner part of the curve be modelled?
3. how should the head be rotated with respect to the rolling of the trunk?

This may require collecting a much larger controlled dataset featuring different curved trajectories, a larger number of walking sequences and more people.

8.2 Recommendations

Due to the technical problems in the Tunnel and the time scale of the PhD period, the largest dataset collected to analyse the performance of matching 3D with 2D gait data contains 50 people. To accurately estimate inter-class variation between subjects, a much larger dataset should be collected. As the number of people in the dataset increases, the processing time substantially increases as each 3D gait cycle in the gallery is projected against all 2D gait cycles in the probe before extracting gait features. An appropriate evaluation protocol should be developed by exploiting modern computer technologies to build a real-time system and reliably estimate the recognition performance. It should be noted that although the size of the collected dataset in the Tunnel is relatively small, the primary aim is to develop 3D to 2D gait matching algorithms for forensic scenarios where there is only one or a small number of video footage available for identity verification. The results obtained by the algorithms showed that matching 3D volumetric data against 2D silhouettes can be done under a controlled conditions and there is no requirement to collect a further dataset of this type.

One of the most surprising things that has come out of this work is the variation in walking style with respect to the trajectory. Clearly, this needs further investigation as neither the large Soton 3D-2D nor the KY4D datasets have enough variations. It is proposed that a series of experiments be designed to investigate this phenomenon. A more sophisticated procedure should be developed by analysing (1) how different body parts (e.g. head, trunk and feet) orientate with respect to the shape of the walking trajectory and (2) the relationship between the curvature of the trajectory and the behaviour of each of these parts. This would help to model the required continuous alignment of the body parts to produce appropriate synthetic gait images according to a target walking trajectory. To achieve this goal, different curved trajectories should be collected with multiple walks on each.

It is highly recommended to enrol people indoors using 3D measurements while a single camera(s) can be set up outdoors to capture gait images of people for recognition. This would help the study of gaits in real world environments. From the acquired experience of collecting the 3D-2D dataset in the Tunnel, it is believed that capturing such dataset will be difficult because the time required by each participant would be much longer. We should also consider recording a range of covariates such as changes in clothing, shoes, carrying conditions, walking speeds, footwear, walking surfaces and time elapsed.

Appendix A

The algorithm of shape from silhouette reconstruction

Algorithm 1 : Shape from Silhouette

Input: array of silhouettes (*Silhouettes*) , array of lookup tables (*lookupTables*).

Output: 3D Volume (*Volume*).

```
for $x = 1 \rightarrow Nx$ do
    for $y = 1 \rightarrow Ny$ do
        for $z = 1 \rightarrow Nz$ do // $x$ , $y$ , $z$ are the axes values in 3D space
            $Counter = 0$
            for $c = 1 \rightarrow NumCams$ do // NumCams is the required number of cameras
                $(u, v) = lookupTables[c, x, y, z]$
                if $Silhouettes[c, u, v] == 1$ then
                    $Counter = Counter + 1$
                else
                    $Break$
                end if
            end for
            if $Counter == Numcams$ then
                $Volume[x, y, z] = 1$
            else
                $Volume[x, y, z] = 0$
            end if
        end for
    end for
end for
```

References

[1] F. Ahmed, P. P. Paul, and M. L. Gavrilova. DTW-based kernel and rank-level fusion for 3D gait recognition using kinect. *The Visual Computer*, 31(6-8):915–924, 2015.

[2] N. Akae, Y. Makihara, and Y. Yagi. The optimal camera arrangement by a performance model for gait recognition. In *IEEE International Conference on Automatic Face Gesture Recognition and Workshops*, pages 292–297, March 2011.

[3] G. Ariyanto and M. Nixon. Model-based 3D gait biometrics. In *International Joint Conference on Biometrics*, October 2011.

[4] K. Bashir, T. Xiang, and S. Gong. Gait recognition using gait entropy image. In *3rd International Conference on Crime Detection and Prevention (ICDP)*, pages 1–6, 2009.

[5] K. Bashir, T. Xiang, and S. Gong. Cross-view gait recognition using correlation strength. In *Proceedings of the British Machine Vision Conference*, pages 109.1–109.11. BMVA Press, 2010.

[6] C. BenAbdelkader, R. Cutler, and L. Davis. Stride and cadence as a biometric in automatic person identification and verification. In *Proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition, FGR '02*, pages 372–377. IEEE Computer Society, Washington, DC, USA, 2002.

[7] 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, Lecture Notes in Computer Science*, volume 2091, pages 284–294. Springer Berlin Heidelberg, 2001.

[8] B. Bhanu and J. Han. Human recognition on combining kinematic and stationary features. In *International Conference on Audio- and Video-Based Biometric Person Authentication, Lecture Notes in Computer Science*, volume 2688, pages 600–608. Springer Berlin Heidelberg, 2003.

[9] F. A. Bobick and Y. A. Johnson. Gait recognition using static activity-specific parameters. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 423–430, 2001.

[10] R. Bodor, A. Drenner, D. Fehr, O. Masoud, and N. Papanikolopoulos. View-independent human motion classification using image-based reconstruction. *Image and Vision Computing*, 27(8):1194 – 1206, 2009.

[11] I. Bouchrika, M. Goffredo, J. Carter, and M. Nixon. On using gait in forensic biometrics. *Journal of forensic sciences*, 56(4):882–889, 2011.

[12] I. Bouchrika and M. S. Nixon. Model-based feature extraction for gait analysis and recognition. In *Computer Vision/Computer Graphics Collaboration Techniques, Lecture Notes in Computer Science*, volume 4418, pages 150–160. Springer Berlin Heidelberg, 2007.

[13] N. Boulgouris, K. Plataniotis, and D. Hatzinakos. Gait recognition using dynamic time warping. In *IEEE 6th workshop on Multimedia Signal Processing*, pages 263–266, 2004.

[14] F. Bukhari and M. N. Dailey. Automatic radial distortion estimation from a single image. *Journal of Mathematical Imaging and Vision*, 45(1):31–45, 2013.

[15] F. M. Castro, M. J. Marín-Jiménez, and R. M. Carnicer. Pyramidal fisher motion for multiview gait recognition. *International Conference on Pattern Recognition*, abs/1403.6950:1692–1697, 2014.

[16] K. Cheung, S. Baker, and T. Kanade. Shape-from-silhouette across time part i: Theory and algorithms. *International Journal of Computer Vision*, 62(3):221–247, 2005.

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

[18] G. Courtine. Tuning of a basic coordination pattern constructs straight-ahead and curved walking in humans. *Journal of Neurophysiology*, 91(4):1524–1535, 2004.

[19] G. Courtine and M. Schieppati. Human walking along a curved path. ii. gait features and emg patterns. *European Journal of Neuroscience*, 18(1):191–205, 2003.

[20] D. Cunado, M. S. Nixon, and J. N. Carter. Using gait as a biometric, via phase-weighted magnitude spectra. In *Audio- and Video-based Biometric Person Authentication, Lecture Notes in Computer Science*, volume 1206, pages 93–102. Springer Berlin Heidelberg, 1997.

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

[22] K. R. Cuntoor, Kale A., A. N. Rajagopalan, N. Cuntoor, and V. Krger. Gait-based recognition of humans using continuous hmms. In *5th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 321–326, 2002.

[23] N. Cuntoor, A. Kale, and R. Chellappa. Combining multiple evidences for gait recognition. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 3, pages 113–116, 2003.

[24] J. E. Cutting and L. T. Kozlowski. Recognizing friends by their walk - gait perception without familiarity cues. *Bulletin of the Psychonomic Society*, 9(5):353–356, 1977.

[25] J. Daugman. Biometric decision landscapes. Technical Report UCAM-CL-TR-482, University of Cambridge, Computer Laboratory, january 2000.

[26] M. Dong, C. Ma, L. Zhu, and N. Lu. A method of distortion correction with parallel lines. *Proceedings of 4th International Symposium on Precision Mechanical Measurements*, 7130:71304G–71304G–6, 2008.

[27] H. El-Alfy, I. Mitsugami, and Y. Yagi. A new gait-based identification method using local gauss maps. In *Computer Vision-ACCV 2014 Workshops*, pages 3–18. Springer, 2014.

[28] J. Flusser, B. Zitova, and T. Suk. *Moments and Moment Invariants in Pattern Recognition*. Wiley Publishing, 2009.

[29] J. P. Foster, M. S. Nixon, and A. Prugel-Bennett. Automatic gait recognition using area-based metrics. *Pattern Recognition Letters*, 24(14):2489 – 2497, 2003.

[30] M. Goffredo, I. Bouchrika, J. N. Carter, and M. S. Nixon. Self-calibrating view-invariant gait biometrics. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 40(4):997–1008, 2010.

[31] R. Gross and J. Shi. The cmu motion of body (MoBo) database. Technical Report CMU-RI-TR-01-18, Robotics Institute, Pittsburgh, PA, June 2001.

[32] Y. Guan, C. Li, and F. Roli. On reducing the effect of covariate factors in gait recognition: a classifier ensemble method. *IEEE transactions on pattern analysis and machine intelligence*, 37(7):1521–1528, 2015.

[33] Y. Guo, G. Xu, and S. Tsuji. Understanding human motion patterns. In *Proceedings of the 12th International Conference on Image Processing and Pattern Recognition (IAPR)*, volume 2, pages 325–329, 1994.

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

[35] J. Han, B. Bhanu, and A. K. Chowdhury. A study on view-insensitive gait recognition. In *ICIP (3)*, pages 297–300, 2005.

[36] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[37] J. B. Hayfron-Acquah, M. S. Nixon, and J. N. Carter. Recognising human and animal movement by symmetry. In *International Conference on Image Processing (ICIP)*, volume 3, pages 290–293, 2001.

[38] T. D. Heseltine. *Face Recognition: Two Dimensional and Three Dimensional Techniques*. PhD thesis, University of York, September 2005.

[39] M. Hofmann, J. Geiger, S. Bachmann, B. Schuller, and G. Rigoll. The {TUM} gait from audio, image and depth (GAID) database: Multimodal recognition of subjects and traits. *Journal of Visual Communication and Image Representation*, 25(1):195 – 206, 2014.

[40] M. Hofmann and G. Rigoll. Exploiting gradient histograms for gait-based person identification. In *the 20th IEEE International Conference on Image Processing (ICIP)*, pages 4171–4175. IEEE, 2013.

[41] H. Iwama, D. Muramatsu, Y. Makihara, and Y. Yagi. Gait verification system for criminal investigation. *IPSJ Transactions on Computer Vision and Applications*, 5(0):163–175, 2013.

[42] 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. *IEEE Transactions on Information Forensics and Security*, 7(5):1511–1521, Oct 2012.

[43] Y. Iwashita and R. Kurazume. Person identification from human walking sequences using affine moment invariants. In *IEEE International Conference on Robotics and Automation (ICRA'09)*, pages 436–441. IEEE, 2009.

[44] Y. Iwashita, K. Ogawara, and R. Kurazume. Identification of people walking along curved trajectories. *Pattern Recognition Letters*, 48:60–69, 2014.

[45] Y. Iwashita, K. Uchino, and R. Kurazume. Gait-based person identification robust to changes in appearance. *Sensors*, 13(6):7884–7901, 2013.

[46] A. Jain, L. Hong, and S. Pankanti. Biometric identification. *Commun. ACM*, 43(2):90–98, February 2000.

[47] A.K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):4–20, 2004.

[48] F. Jean, R. Bergevin, and A. Branzan-Albu. Trajectory normalization for viewpoint invariant gait recognition. In *19th International Conference on Pattern Recognition (IAPR)*, December 2008.

[49] G. Johansson. *Visual Motion Perception*. Scientific American offprints. Scientific American Incorporated, 1975.

[50] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001.

[51] A. Kale, A. K. Chowdhury, and R. Chellappa. Towards a view invariant gait recognition algorithm. In *Proceedings of IEEE Conference on Advanced Video and Signal Based Surveillance*, pages 143–150, 2003.

[52] A. Kale, N. Cuntoor, B. Yegnanarayana, A. Rajagopalan, and R. Chellappa. Gait analysis for human identification. In *Audio-and Video-Based Biometric Person Authentication*, pages 1058–1058. Springer, 2003.

[53] A. Kale, A. Sundaresan, A. Rajagopalan, N. Cuntoor, A. Roy-Chowdhury, V. Krüger, and R. Chellappa. Identification of humans using gait. *IEEE Transactions on Image Processing*, 13(9):1163–1173, 2004.

[54] A. Kharb, V. Saini, Y. Jain, and S. Dhiman. A review of gait cycle and its parameters. *International Journal of Computational Engineering and Management (IJCEM)*, 13:78–83, 2011.

[55] J. Kruger and R. Westermann. Acceleration techniques for GPU-based volume rendering. In *Proceedings of the 14th IEEE Visualization (VIS'03)*, Washington, DC, USA, 2003.

[56] W. Kusakunniran, Q. Wu, H. Li, and J. Zhang. Multiple views gait recognition using view transformation model based on optimized gait energy image. In *IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 1058–1064, 2009.

[57] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li. Support vector regression for multi-view gait recognition based on local motion feature selection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 974–981. IEEE, 2010.

[58] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li. Gait recognition under various viewing angles based on correlated motion regression. *IEEE Trans. Circuits Syst. Video Techn.*, 22(6):966–980, 2012.

[59] W. Kusakunniran, Q. Wu, J. Zhang, H. Li, and L. Wang. Recognizing gaits across views through correlated motion co-clustering. *IEEE Transactions on Image Processing*, 23(2):696–709, 2014.

[60] H. W. Lam, K. H. Cheung, and N. K. Liu. Gait flow image: A silhouette-based gait representation for human identification. *Pattern Recognition*, 44(4):973 – 987, 2011.

[61] T. H. Lam and R. S. Lee. A new representation for human gait recognition: Motion silhouettes image (msi). In *Proceedings of the international conference on Advances in Biometrics*, number 7, pages 612–618. Springer-Verlag, Berlin, Heidelberg, 2006.

[62] A. Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(2):150–162, 1994.

[63] L. Lee. Gait analysis for classification. Technical report, Massachusetts Institute of Technology, 2003.

[64] L. Lee and W. E. Grimson. Gait analysis for recognition and classification. In *Proceedings of 5th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 148–155, 2002.

[65] S. Lee, Y. Liu, and R. T. Collins. Shape variation-based frieze pattern for robust gait recognition. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Minneapolis, Minnesota, USA*, pages 1–8, 2007.

[66] S. Z. Li. *Encyclopedia of Biometrics, 1st edition.* Springer Publishing Company, Incorporated, 2009.

[67] J. Little and J. Boyd. Recognizing people by their gait: the shape of motion. *Videre: Journal of Computer Vision Research*, 1(2):1–32, 1998.

[68] J. Liu and N. Zheng. Gait history image: A novel temporal template for gait recognition. In *IEEE International Conference on Multimedia and Expo.*, pages 663–666, 2007.

[69] N. Liu and Y. Tan. View invariant gait recognition. In *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pages 1410–1413, 2010.

[70] Y. Liu, R. T. Collins, and Y. Tsin. Gait sequence analysis using frieze patterns. In *Proceedings of European Conference on Computer Vision*, pages 657–671, 2001.

[71] Z. Liu, L. Malave, A. Osuntogun, P. Sudhakar, and S. Sarkar. Toward understanding the limits of gait recognition. *Proceedings of SPIE Biometric Technology for Human Identification*, 5404(195):195–205, 2004.

[72] Z. Liu and S. Sarkar. Simplest representation yet for gait recognition: Averaged silhouette. In *Proceedings of the 17th International Conference on Pattern Recognition (ICPR)*, volume 4, pages 211–214, 2004.

[73] D. López-Fernández, F. Madrid-Cuevas, A. Carmona-Poyato, M. Marín-Jiménez, R. Munoz-Salinas, and R. Medina-Carnicer. Viewpoint-independent gait recognition through morphological descriptions of 3D human reconstructions. *Image and Vision Computing*, 48:1–13, 2016.

[74] D. López-Fernández, F. Madrid-Cuevas, A. Carmona-Poyato, R. Muñoz-Salinas, and R. Medina-Carnicer. A new approach for multi-view gait recognition on unconstrained paths. *Journal of Visual Communication and Image Representation*, 38:396–406, 2016.

[75] D. López-Fernández, F. J. Madrid-Cuevas, Á Carmona-Poyato, M. J. Marín-Jimnez, and R. Muñoz Salinas. The AVA multi-view dataset for gait recognition. In *Activity Monitoring by Multiple Distributed Sensing*, Lecture Notes in Computer Science, pages 26–39. Springer International Publishing, 2014.

[76] D. López-Fernández, F. J. Madrid-Cuevas, A. Carmona-Poyato, R. Muñoz-Salinas, and R. Medina-Carnicer. Multi-view gait recognition on curved trajectories. *Proceedings*

of the 9th International Conference on Distributed Smart Camera - ICDSC '15, pages 116–121, 2015.

[77] D. Lpez-Fernndez, F.J. Madrid-Cuevas, A. Carmona-Poyato, R. Muoz-Salinas, and R. Medina-Carnicer. Entropy volumes for viewpoint-independent gait recognition. *Machine Vision and Applications*, pages 1–16, 2015.

[78] N. Lynnerup and P. Larsen. Gait as evidence. *IET Biometrics*, 3(2):47–54, 2014.

[79] Y. Makihara, T. Kimura, F. Okura, I. Mitsugami, M. Niwa, C. Aoki, A. Suzuki, D. Muramatsu, and Y. Yagi. Gait collector: An automatic gait data collection system in conjunction with an experience-based long-run exhibition. In *International Conference on Biometrics (ICB)*, pages 1–8, 2016.

[80] Y. Makihara, H. Mannami, A. Tsuji, M. Hossain, K. Sugiura, A. Mori, and Y. Yagi. The OU-ISIR gait database comprising the treadmill dataset. *IPSJ Transactions on Computer Vision and Applications*, 4:53–62, 2012.

[81] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi. Gait recognition using a view transformation model in the frequency domain. In *Proceedings of the 9th European conference on Computer Vision - Volume Part III, ECCV'06*, pages 151–163. Springer-Verlag, Berlin, Heidelberg, Graz, Austria, 2006.

[82] D. S. Matovski, M. S. Nixon, S. Mahmoodi, and J. N. Carter. The effect of time on gait recognition performance. *IEEE Transactions on Information Forensics and Security*, 7(2):543–552, 2012.

[83] L. Middleton, D.K. Wagg, A.I. Bazin, J.N. Carter, and M.S. Nixon. A smart environment for biometric capture. In *IEEE International Conference on Automation Science and Engineering (CASE) '06.*, pages 57–62, October 2006.

[84] D. Muramatsu, A. Shiraishi, Y. Makihara, M.Z. Uddin, and Y. Yagi. Gait-based person recognition using arbitrary view transformation model. *IEEE Transactions on Image Processing*, 24(1):140–154, January 2015.

[85] P. M. Murray, B. A. Drought, and R. C. Kory. Walking patterns of normal men. *Journal of Bone and Joint Surgery*, 46(2):335–360, March 1964.

[86] H. Nakajima, I. Mitsugami, and Y. Yagi. Depth-based gait feature representation. *IPSJ Transactions on Computer Vision and Applications*, 5:94–98, July 2013.

[87] J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, January 1965.

[88] M Nixon, J. Carter, J. Shutler, and M. Grant. New advances in automatic gait recognition. *Information Security Technical Report*, 7(4):23–35, 2002.

[89] M. S. Nixon and J. N. Carter. Advances in automatic gait recognition. In *IEEE Face and Gesture Analysis, FG04*, pages 11–16. IEEE CS Press, 2004.

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

[91] S. A. Niyogi and E. H. Adelson. Analyzing and recognizing walking figures in XYT. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 469–474, 1994.

[92] C. Padole and H. Proença. An aperiodic feature representation for gait recognition in cross-view scenarios for unconstrained biometrics. *Pattern Analysis and Applications*, pages 1–14, 2015.

[93] N. K Pandey, W. H Abdulla, and Z. Salcic. Gait recognition using sub-vector quantisation technique. *International Journal of Machine Intelligence and Sensory Signal Processing*, 1(1):68–90, 2013.

[94] P. Phillips, H. Moon, S. Rizvi, and P. Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on pattern analysis and machine intelligence*, 22(10):1090–1104, 2000.

[95] P. J. Phillips, S. Sarkar, I. Robledo, P. Grother, and K. Bowyer. The gait identification challenge problem: Data sets and baseline algorithm. In *Proceedings of 16th International Conference on Pattern Recognition*, volume 1, pages 385–388, 2002.

[96] S. Salvador and P. Chan. Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis*, 11(5):561–580, October 2007.

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

[98] R. D. Seely. *On a three-dimensional gait recognition system*. PhD thesis, University of Southampton, July 2010.

[99] R. D. Seely, M. Goffredo, J. N. Carter, and M. S. Nixon. View invariant gait recognition. In *Handbook of Remote Biometrics, Advances in Pattern Recognition*, pages 61–81. Springer London, 2009.

[100] R. D. Seely, S. Samangooei, M. Lee, J. N. Carter, and M. S. Nixon. The university of southampton multi-biometric tunnel and introducing a novel 3D gait dataset. In *2nd IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–6, 2008.

[101] Y. Shiqi, T. Daoliang, and T. Tieniu. Modelling the effect of view angle variation on appearance-based gait recognition. In *Computer Vision ACCV 2006, Lecture Notes in Computer Science, Springer Berlin Heidelberg*, volume 3851, pages 807–816. 2006.

[102] J. Shutler, M. Grant, M. S. Nixon, and J. N. Carter. On a large sequence-based human gait database. In *Proc. RASC*, pages 66–72. Springer Verlag, 2002.

[103] J. D. Shutler and M. S. Nixon. Zernike velocity moments for sequence-based description of moving features. *Image and Vision Computing*, 24(4):343 – 356, 2006.

[104] J. D. Shutler, M. S. Nixon, and C. J. Harris. Statistical gait recognition via velocity moments. In *IEEE Colloquium: Visual Biometrics*, number 00/018, pages 11/1–11/5, 2000.

[105] S. Sivapalan, D. Chen, S. Denman, S. Sridharan, and C. Fookes. 3D ellipsoid fitting for multi-view gait recognition. In *Proceedings of 8th IEEE International Conference on Advanced Video and Signal Based Surveillance*, pages 355–360. IEEE Computer Society, Washington, DC, USA, 2011.

[106] S. Sivapalan, D. Chen, S. Denman, S. Sridharan, and C. Fookes. Gait energy volumes and frontal gait recognition using depth images. In *International Joint Conference on Biometrics*, pages 1–6, 2011.

[107] J. E. Solem. *Programming Computer Vision with Python - Tools and algorithms for analyzing images.* O'Reilly, 2012.

[108] M. Sreenivasa, I. Frissen, J. Souman, and M. Ernst. Walking along curved paths of different angles: the relationship between head and trunk turning. *Experimental Brain Research*, 191(3):313–320, 2008.

[109] A. Sundaresan, R. Roy-Chowdhury, and R. Chellappa. A hidden markov model based framework for recognition of humans from gait sequences. In *Proceedings of International Conference on Image Processing (ICIP)*, volume 2, pages II–93–6, 2003.

[110] D. Tan, K. Huang, S. Yu, and T. Tan. Efficient night gait recognition based on template matching. In *18th International Conference on Pattern Recognition*, volume 3, pages 1000–1003, 2006.

[111] D. Tan, K. Huang, S. Yu, and T. Tan. Uniprojective features for gait recognition. In *Advances in Biometrics*, pages 673–682. Springer, 2007.

[112] R. Tanawongsuwan and A. Bobick. Gait recognition from time-normalized joint-angle trajectories in the walking plane. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages II–726, 2001.

[113] J. Tang, J. Luo, T. Tjahjadi, and Y. Gao. 2.5D multi-view gait recognition based on point cloud registration. *Sensors*, 14(4):6124–6143, 2014.

[114] E. Tassone, G. West, and S. Venkatesh. Temporal PDMs for gait classification. In *Proceedings of 16th International Conference on Pattern Recognition*, volume 2, pages 1065–1068, 2002.

[115] G. V. Veres, L. Gordon, J. N. Carter, and M. S. Nixon. What image information is important in silhouette-based gait recognition? In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages II–776–II–782, 2004.

[116] D. K. Wagg and M. S. Nixon. On automated model-based extraction and analysis of gait. In *6th International Conference on Automatic Face and Gesture Recognition*, pages 11–16. IEEE Computer Society Press, 2004.

[117] L. Wang, H. Ning, W. Hu, and T. Tan. Gait recognition based on procrustes shape analysis. In *Proceedings of International Conference on Image Processing*, volume 3, pages III–433–III–436, 2002.

[118] L. Wang, H. Ning, T. Tan, and W. Hu. Fusion of static and dynamic body biometrics for gait recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(2):149–158, 2004.

[119] L. Wang, T. Tan, H. Ning, and W. Hu. Silhouette analysis-based gait recognition for human identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1505–1518, 2003.

[120] D. A. Winter. *Biomechanics and Motor Control of Human Movement*. Wiley, 3rd edition, August 2004.

[121] X. Xing, K. Wang, T. Yan, and Z. Lv. Complete canonical correlation analysis with application to multi-view gait recognition. *Pattern Recognition*, 50:107–117, 2016.

[122] C. Y. Yam, M. S. Nixon, and J. N. Carter. Automated person recognition by walking and running via model-based approaches. *Pattern Recognition*, 37(5):1057 – 1072, 2004.

[123] S. Yu, D. Tan, and T. Tan. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *18th International Conference on Pattern Recognition*, volume 4, pages 441–444, 2006.

[124] D. Zhang and G. Lu. Shape-based image retrieval using generic fourier descriptor. *Signal Processing: Image Communication*, 17(10):825 – 848, 2002.

[125] G. Zhao, G. Liu, H. Li, and M. Pietikainen. 3D gait recognition using multiple cameras. In *7th International Conference on Automatic Face and Gesture Recognition (FGR)*, pages 529–534, 2006.