

## University of Southampton Research Repository ePrints Soton

Copyright © and Moral Rights for this thesis are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given e.g.

AUTHOR (year of submission) "Full thesis title", University of Southampton, name of the University School or Department, PhD Thesis, pagination

UNIVERSITY OF SOUTHAMPTON

# Techniques for orientation independent gait analysis

by

Robert Trevor Boston

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

in the

Faculty of Engineering, Science and Mathematics  
School of Electronics and Computer Science

October 2008

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

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

Doctor of Philosophy

by Robert Trevor Boston

Gait recognition algorithms are being increasingly widely researched, however a common assumption is that the subject will be presented side on to the camera. In practice it may not be possible to capture data from this view, so a useful gait recognition algorithm will have to provide a measure of orientation independence.

Three gait recognition algorithms are examined and found to perform poorly with non-normal orientation. The complex detail used for recognition can not be translated between orientations in a holistic silhouette manner.

It is shown that orientation independent features can be extracted using a human model. The algorithm is developed and tested on live captured data and found to perform better across orientations than silhouette based approaches. The performance recorded at a single orientation is lower than that of other approaches, however only the motion of the subject is currently used for recognition. More accurate motion estimation will increase performance as will the inclusion of other model based features.

# Contents

<b>Acknowledgements</b>	<b>x</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Context and Contribution</b>	<b>3</b>
2.1 Biometrics	3
2.2 Typical Biometric Process	5
2.2.1 Different Biometrics	7
2.3 Gait Recognition	9
2.3.1 Model Based Gait Recognition	11
2.3.2 Silhouette Based Gait Recognition	13
2.4 Orientation Independence	13
2.5 Contribution.	17
<b>3 Baseline Algorithm at Multiple Orientations</b>	<b>19</b>
3.1 Southampton GTRI Artificial Walker	19
3.2 Calculating the Period	21
3.3 Original Algorithm	23
3.4 Modified Version	24
3.5 Experimental Setup.	25
3.6 Results	27
3.7 Conclusions	32
<b>4 Average Silhouette at Multiple Orientations</b>	<b>33</b>
4.1 The Average Silhouette Algorithm	34
4.2 Experimentation upon SGAW Data	36
4.3 Recognition Experiments with the SGAW	40
4.3.1 Gallery at Probe Orientation	41
4.4 Recognition Rates with the SGAW data using Planar People	45
4.5 Recognition with real subjects	49
4.6 Conclusions	49
<b>5 Orientation Independent Properties of Moments.</b>	<b>52</b>
5.1 Moment Generation	52
5.2 Analysis of moment variance.	54
5.3 Identification of Subject	55
5.4 Determining the Orientation	58
5.5 Conclusions from Moment Experiments	60

---

<b>6</b>	<b>Orientation Independent Model Based Recognition</b>	<b>65</b>
6.1	The Model . . . . .	66
6.2	Silhouette Generation . . . . .	69
6.3	Silhouette Alignment Optimisation . . . . .	71
6.4	Improved initialisation . . . . .	74
6.5	Recognition . . . . .	75
6.6	Discussion and Conclusions . . . . .	78
<b>7</b>	<b>Improved Model Based Algorithm</b>	<b>79</b>
7.1	Improved Fitness Metric . . . . .	79
7.2	Using head ankle angle . . . . .	80
7.3	Removing hip constraint . . . . .	81
7.4	Template based lower legs . . . . .	83
7.5	Results . . . . .	87
7.6	Discussion and Conclusions . . . . .	87
<b>8</b>	<b>Future Work and Conclusions</b>	<b>91</b>
8.1	Conclusions. . . . .	91
8.2	Future work . . . . .	92
<b>A</b>	<b>Baseline Algorithm Results</b>	<b>94</b>
A.1	Within Orientation Confusion Matrices . . . . .	94
A.2	Between Orientation Confusion Matrices . . . . .	96
<b>B</b>	<b>Average Silhouette Algorithm Results</b>	<b>98</b>
B.1	Within Orientation Confusion Matrices . . . . .	98
B.2	Between Orientation Confusion Matrices . . . . .	100
	<b>Bibliography</b>	<b>102</b>

# List of Figures

2.1	How enrolment, identification and verification of users relate to each other.	5
2.2	A confusion matrix, the dark diagonal 3 by 3 squares show the relatively large inter-subject to intra-subject variance.	6
2.3	A heelstrike (a), double support (b), thighs occluding (c) and lower leg occlusion.	10
2.4	A common layout for a gait recognition experiment.	11
2.5	Given the track angle $a$ to the image plane and a ray from the subject at angle $b$ to the camera's view axis the orientation of the subject is $a - b$ .	14
2.6	The orientation of a subject as they cross the camera's field of view	15
2.7	The horizontal rescaling of a subject as they cross the camera's field of view.	16
3.1	The motion capture points and the construction of the 3D model.	20
3.2	Some examples of the data that can be produced using the 3D model.	21
3.3	Measured widths of bounding boxes of a silhouette sequence.	22
3.4	Measured widths, an initialised curve and the final fitted curve.	24
3.5	Encoding a flat binary image.	26
3.6	The confusion matrix of all runs of all subjects at 0,45 and 90 degrees. 4 by 4 blocks represent a comparison of 4 instances of one subject with 4 instances of another. Dark 4 by 4 boxes along the diagonal indicate there is little difference between instances of the same subject and a large difference between subjects.	28
3.7	The confusion matrix of all runs of all subjects at 0 degrees vs other orientations.	28
3.8	The recognition rate at each orientation, where, orientation of the probe = orientation of the gallery	30
3.9	The recognition rate with galleries and probes at different orientations.	30
4.1	3 Sample frames from person 5105, instance 1, angle = 0, with the sequence average	36
4.2	Example feature vectors. 4 instances of the same person (a-d), 4 different people (5112,5114,5115,5128) e-h and 4 different orientations (0,5,30,90) i-l.	37
4.3	For orientation 0, the difference between the 4 instances and mean of each subject. The mean deviation of each person is similar but there is a significant variance.	38
4.4	The differences of the 4 instances of subject 5105 from the mean of 5105 at 5 degree intervals to 90 degrees.	38
4.5	The difference of each subject from the global mean feature vector.	39

4.6	The confusion matrix of all runs of all subjects at 0,45 and 90 degrees. Each pixel row and column represents one instance of a subject. The 4 by 4 boxes of a similar shade show the internal consistency of each subject. The 4 by 4 dark boxes along the diagonal indicates there is little difference between instances of the same subject and a large difference between subjects. . . . .	39
4.7	The confusion matrix of all runs of all subjects at 0 degrees vs other orientations. . . . .	40
4.8	The mean difference between adjacent angles . . . . .	40
4.9	The mean difference between 0 degrees and the other angles . . . . .	41
4.10	The mean difference between all angles . . . . .	41
4.11	The recognition rate at each orientation, where, orientation of the probe = orientation of the gallery . . . . .	42
4.12	The recognition rate with galleries and probes at different orientations. . . . .	43
4.13	The recognition rate attainable if only galleries at 0,45 and 90 are available and probes are compared to the closest gallery. There are large drops in recognition when the the probe is far from one of the three galleries. . . . .	45
4.14	The recognition rate accross all orientaions with a gallery at 5 degrees. . . . .	45
4.15	The recognition rate with galleries and probes at different orientations; (a) with a planar person approximation, (b) without any correction. . . . .	48
4.16	Example average silhouettes. . . . .	50
5.1	Moment (2,0) of frame 8 over ninety degrees plotted for 10 subjects. The moment appears to change with person identity but more with orientation. . . . .	54
5.2	When only 3 galleries at 0,45 and 90 degrees are used, probes from intermediate orientations have a reduced recognition rate when compared to one of the 3 galleries. If a pair of galleries is interpolated to generate a virtual gallery at the same orientation as the probe, recognition at unrepresented orientations is improved. . . . .	56
5.3	Average recognition rates are reduced when the spacing between the galleries is increased. Interpolating between galleries marginally improves recognition. Note that the probes are spaced at 5 degree intervals so the first point represents having a recoded gallery at the orientation of each probe. . . . .	56
5.4	Five subjects were used to generate average moments for each angle and five were used for the experiment. Normalising the moments of each person using the averages gave a marginal improvement in recognition. . . . .	57
5.5	The recognition rate between probes and galleries at all orientations. . . . .	58
5.6	The similarity of moments from different orientations. The sum of the distance between every run of every subject between all combinations of orientation. Each square shows the sum of the distances of all runs rescaled to a greyscale. White is the minimum and black the maximum. . . . .	59
5.7	Only 50% of the orientations were recognised correctly, however 92% were estimated within 5 degrees of the real orientation. . . . .	59
5.8	Two simple shapes (a and c) and outer edges projected onto a plane as they are rotated about their cross (b and d), 0° is horizontal. . . . .	63
5.9	The projected outer edge positions (b) as the shape (a) is rotated about the cross, 0° is horizontal. . . . .	63

5.10	Two shapes (a and c) and outer edges projected onto a plane as they are rotated about their cross (b and d), $0^\circ$ is horizontal. . . . .	64
6.1	Three example silhouettes generated from the model. . . . .	66
6.2	The parameters used to generate a silhouette. . . . .	70
6.3	Real (a) and artificial (b) silhouettes with their overlap (c), white is background, grey is the overlap, black is the difference between the two. . . . .	72
6.4	The speed and depth of the several different optimization techniques. . . . .	73
6.5	Extracted thigh and shin motion. . . . .	74
6.6	Extracted thigh and shin motion with improved initialisation. . . . .	75
6.7	Motion extracted from more than one stride (a), and phase shifted by $\pi$ (b). . . . .	76
6.8	Motion extracted from more than one stride (a), left leg shifted by $\pi$ , and wrapped to one period (b). . . . .	77
7.1	The real silhouette used in comparison (a) and the corresponding distance map (b). . . . .	80
7.2	The new head ankle measurements in relation to the existing parameters. . . . .	81
7.3	Two images of aligned synthetic thighs, the x coordinate of the top is fixed, the length and deflection from vertical of the left rectangle is optimised with a brute force algorithm to minimise the difference. . . . .	82
7.4	The fitness and optimising values of angle and length at fixed horizontal displacements. . . . .	82
7.5	The optimising values of angle and thigh plotted against each other under various horizontal offsets. . . . .	82
7.6	The first process is the generation of a template based model. Lower legs are segmented out of the original silhouettes. . . . .	84
7.7	(a) and (b) segmented legs, the initial template (c) and the final template (d). . . . .	84
7.8	The second process in generating a template based model. Segmented legs are used to construct an average lower leg. . . . .	84
7.9	A head template created similarly to the leg templates. . . . .	85
7.10	2 example silhouettes generated from the template model. . . . .	86
7.11	Rescaling the shin for different shin lengths. . . . .	86
7.12	2 example silhouettes generated from the template model. . . . .	86
7.13	2 example silhouettes generated from the template model. . . . .	87
7.14	Captured images and the silhouettes calculated. (a) and (b) an image captured from the $0^\circ$ . (c) and (d), interlaced image with noticeable vertical blur to the lower leg, (e) and (f), blurring increasing the occlusion of the legs, (g) and (h) horizontal blurring of the lower leg. . . . .	89
7.15	examples of the same foot segmented from different parts of an interlaced sequence. (a) is stationary with minimal corruption, (b) has vertical blur, (c) and (d) have horizontal blur. . . . .	90
A.1	The confusion matrix of all runs of all subjects at $0, 5, \dots, 90$ degrees. 4 by 4 dark boxes along the diagonal indicate there is little difference between instances of the same subject, white areas show large difference between subjects. . . . .	94
A.2	Figure A.1 continued . . . . .	95

---

A.3	Figure A.1 continued . . . . .	95
A.4	Confusion matrices of feature vectors recorded at 0 degrees vs other orientations. . . . .	96
A.5	Figure A.4 continued. . . . .	97
B.1	The confusion matrix of all runs of all subjects at 0,5,...,90 degrees. 4 by 4 dark boxes along the diagonal indicate there is little difference between instances of the same subject, white areas show large difference between subjects. . . . .	98
B.2	Figure B.1 continued . . . . .	99
B.3	Figure B.1 continued . . . . .	99
B.4	Confusion matrices of feature vectors recorded at 0 degrees vs other orientations. . . . .	100
B.5	Figure B.4 continued. . . . .	101

# List of Tables

2.1	Imaginary data from a beard recognition system. . . . .	7
2.2	A comparison of the performance of several biometrics. Adapted from Lang and Leopold (1997) . . . . .	9
3.1	The heuristics controlling the width of body parts. . . . .	20
3.2	Recognition rates, when the probe and gallery are at the same orientation.	29
3.3	Recognition rates for when the probe and gallery are at different orienta- tions. . . . .	31
4.1	Recognition rates for when the probe and gallery are at the same orientation.	42
4.2	Recognition rates when the probe and gallery are at different orientations.	44
4.3	Recognition rates for when the probe and gallery are at different orienta- tions but corrected with the planar people approximation. . . . .	47
4.4	Recognition rates between orientations. . . . .	49
6.1	Angles required to fully pose the body in 3D. . . . .	67
6.2	Parameters used to draw a silhouette. . . . .	70
6.3	Recognition rates with the model based algorithm. . . . .	78
7.1	Recognition rates using the new template fitting algorithm between ori- entations. . . . .	87

## DECLARATION OF AUTHORSHIP

I,

declare that the thesis entitled "Techniques for orientation independent gait analysis" 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:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- none of this work has been published before submission.

Signed:

Date:

## **Acknowledgements**

I want to first say thank you to my family, Mum, Dad, Dave, Auntie Winnie and Toby for their help and unwavering support and patience on this long and difficult journey. Thanks Zoe and SJ for your faith and for keeping me going through the tough bits, and thank you John for your excellent guidance and patience.

# Chapter 1

## Introduction

Traditional forms of identification are based upon secret knowledge or unique physical objects. Secret knowledge in the form of passwords or personal identification numbers can be forgotten or learnt by an untrusted third party. Unique objects such as keys or cards can be lost, copied or stolen. Both forms of identification require effort and co-operation by a user.

Biometric technologies offer a potential third method of identification with several advantages compared to knowledge and object based systems. Biometrics are properties of people that can be used to recognise them, such as fingerprints or facial features. By using something the user is rather than something they know or possess it makes it increasingly difficult for an impostor to steal a legitimate user's identity. Biometrics can be extracted with little, if any effort by a user, making covert identification possible. It is the increases in computing power and the advancement of computer vision over the last decade that has increased the practicality of automated biometric systems.

Humans have the ability to recognise each other by their gait so the aim of gait based biometric research is to give this ability to an automated system. Gait is appealing as it can be measured at a distance where other biometrics are ineffective. For example, using low resolution cameras where the subjects face is a few pixels wide, face recognition is impossible. There are many potential applications of such a system. These could include replacing traditional identification for access to restricted areas, reconciling people between distinct camera footage, or any situation where the identity of a person only measurable at a distance is of some value.

As gait recognition is a relatively young area of research, most common approaches have been developed under simplifying assumptions. These assumptions include flat terrain, constant walking speed and simple clothing. One common assumption is that the subject is presented normal to the camera, i.e., travelling perpendicular to the camera's axis. In real world scenarios this situation is unlikely.

Under lab conditions, when a subject is travelling in a path perpendicular to the camera's view axis, a normal view is presented to the camera only at the instant they are at the image centre. In real world conditions, normal views will be rare apart from in highly controlled environments, and then only for a limited number of cameras in the optimal positions can be used.

A system that can recognise subjects at a wide range of orientations can make use of data captured at non-normal orientations, in less tightly controlled environments, such as in airports as opposed to corridors.

Body parameters such as height, stride length and limb motion can be translated to a normal view using motion captured data from a reference subject [Johnson and Bobick (2001), Carter and Nixon (2000b), Spencer and Carter (2002)]. It is possible to translate the silhouette to a normal orientation before measurements are taken (Chowdhury et al., 2003) but limiting assumptions have to be made. Visual hulls can be constructed and silhouettes projected to normal views (Lee, 2001) but this requires a large number of cameras.

In chapter 2 a summary of gait recognition is presented, introducing some of the different approaches taken. The chapter finishes by showing why orientation independent gait recognition is desirable and the current approaches. Next, in chapter 3 the performance of a baseline algorithm is analysed over a wide range of orientations with the use of synthetic data. Following on in chapter 4, the average silhouette algorithm is shown to have marginally better performance than the baseline. The results of the synthetic experimentation are verified on live captured data. Chapter 5 performs a detailed analysis of a moment based algorithm, attempting to find properties that are orientation independent. The chapter finishes with the conclusion that there is no transformation that can be applied to a silhouette or a derivative thereof without the use of a model to guide it. Chapter 6 extends a model based algorithm into 3d and extracts orientation independent properties that can be used for recognition. Its performance however is poor compared to the silhouette based methods of the previous 3 chapters. Chapter 7 modifies the algorithm described in chapter 6 and achieves orientation independence comparable to the average silhouette algorithm based purely on gait motion and with much room for improvement. Finally, chapter 8 concludes this work and points out several future avenues of investigation.

## Chapter 2

# Context and Contribution

Traditional identification methods use unique physical objects or secret knowledge to ascertain the identity of an individual and restrict access to resources. However, physical objects such as keys or cards can be lost or stolen. Secret knowledge in the form of PINs or passwords can be learnt by untrusted third parties. Both forms of identification require effort by a user and can be withheld.

Biometrics can be used as an alternative or in addition to objects and knowledge. By using a property of the subject as the ‘key’, transferring the key to someone else is more difficult. As computing power increases and the computer vision field matures, biometric technology is emerging that can identify a subject in a range of previously impractical situations. By combining biometrics with both traditional forms of identification, increasingly secure and reliable identification systems can be developed.

### 2.1 Biometrics

The word biometric is derived from ”bios” and ”metron”, the Greek for life and degree. Biometrics are, most generally, statistics pertaining to biological phenomenon, however, the word is most commonly used to refer to the science, technology and measurements used for human identification.

Most current forms of identification rely on private possessions (something the user owns) or private knowledge (something the user knows). However, biometric identification relies upon a characteristic measurable property that can be used to discriminate between people (something the user is). Knowledge, physical objects and biometric forms of identification have weaknesses, however by combining two or more modes of identification, increasingly robust security systems can be implemented.

Biometrics can not be lost or stolen in the same manner that possessions can. It is possible that a biometric would cease working if the physical property changed significantly, for example a facial injury may cause the subject to appear different enough to cause a face recognition system to incorrectly identify a subject as an impostor. This sort of lost biometric scenario is similar to forgetting private information and although inconvenient it may be an extremely rare occurrence depending on the biometric. Also, just as private knowledge can not be 'found' by someone else, neither can a biometric.

Private knowledge in theory should be more secure than possessions. However, with the increased use of passwords and PINs for daily functions people are reaching data overload and resort to recording private information or using the same password for several systems. Once the password is committed to paper or entered into a phone they become as vulnerable to loss as a possession. A survey in 2003 found 30% of people questioned kept a written copy of their computer password, thus making it possible for an impostor to steal the private knowledge. It is also possible to trick a legitimate user into revealing private information.

It is impossible to share biometric information for identification purposes whereas, it is easy to share property and easier to share information. Although sharing identifiers can be a useful, simple and safe practice in some situations, it can be a vulnerability in others.

Some biometrics can be captured covertly and without the subjects knowledge or assistance, this makes it harder for a subject to withhold biometric information than other forms of identification. This type of biometric identification is common in forensic applications where criminals are more likely to leave fingerprints than a driving license.

Most biometric based identification applications require users to register with the system prior to authentication. Whether initializing a traditional system with the distribution of physical ID or knowledge is easier depends on the application, it may not be practical or possible.

One of the most attractive aspects of biometric technology is the ability to automate a system. Automated biometric based systems have the ability to identify people with significantly less effort and interruption to their normal activities, this makes it possible to identify a large number of people quicker. Applications include scanning crowds for people or verifying a person's identity as they log on a computer.

There are some concerns about identity theft and biometrics technology. If a password or key is stolen it is possible to issue a new password or change the locks, if a biometric is accurately captured and an impostor is able to present it convincingly there is no analogous way to issue a new biometric. The potential also exists for the invasion of privacy; some biometrics may expose personal medical data or allow the automated tracking and surveillance of law abiding citizens.

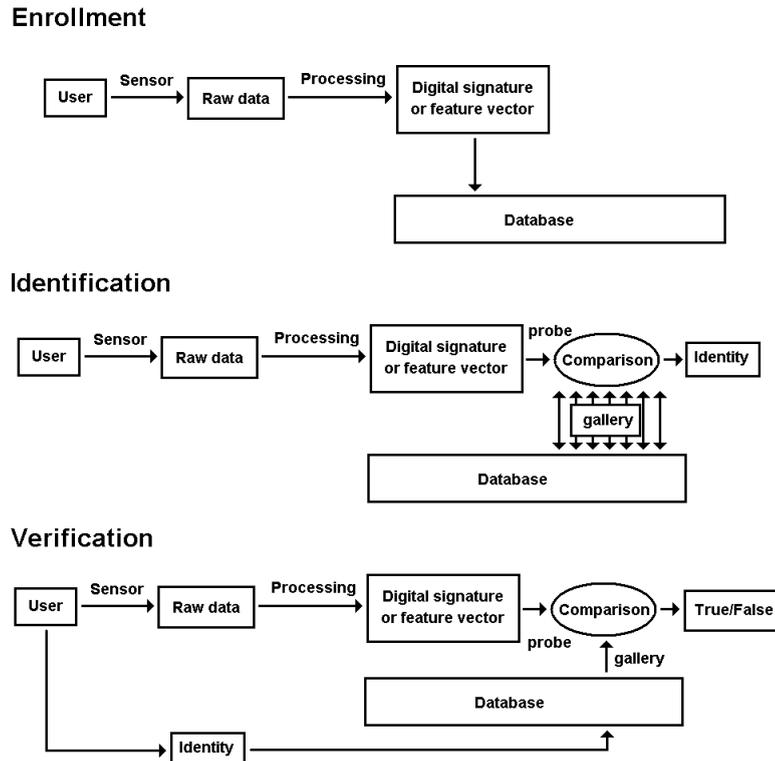


FIGURE 2.1: How enrolment, identification and verification of users relate to each other.

## 2.2 Typical Biometric Process

The goal of a biometric system is to ascertain the identity of a subject. The first stage in a biometric system is to capture raw data such as a photo of someone's face, or a recorded speech segment. Typically this data is processed to extract biometric information, remove noise and highlight unique characteristics, for example removing shadows on a subject's face or measuring the distance between the eyes. During enrolment the processed data is stored in a database, for identification or verification the processed data is compared to those stored in the database to identify the subject or verify they are who they claim.

Figure 2.1 shows how enrolment, identification and verification are related. All three tasks take raw data captured by the sensor and apply some processing to it resulting in a feature vector. The processing typically has several functions; noise can be reduced and external factors such as lighting and pose can be compensated for, also the representation can be compressed considerably. In a fingerprint based system the raw data might be a raw image of the finger, the processing could involve extracting minutiae, reducing the size of the data from an image to information consisting of 30-40 points.

For identification tasks it is necessary to compare the feature vector generated from a subject with the feature vectors stored in the database, the first is referred to as the probe

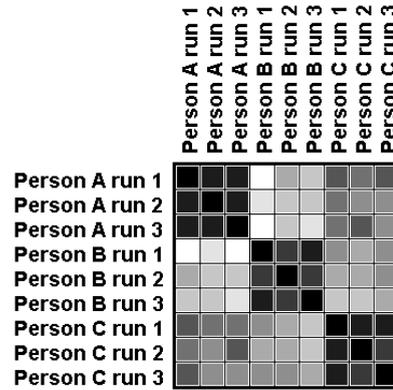


FIGURE 2.2: A confusion matrix, the dark diagonal 3 by 3 squares show the relatively large inter-subject to intra-subject variance.

and the database as the gallery. A common method of comparing feature vectors and deciding identity is with the ‘K nearest neighbour’ algorithm. Biometric technologies are commonly used to identify and to verify the identity of a subject. Identification involves calculating which of a database of people the subject is.

In a verification scenario, the similarity of two feature vectors is calculated. If the distance between them is sufficiently small, a decision is made that the user is who they claim to be, otherwise they are rejected. The rejection of a legitimate user (a false reject) is a type 1 error, and the acceptance of an impostor (false accept) is a type 2 error. The ratio of false rejects and false accepts to the total number of trials gives the false reject rate (FRR) and the false accept rate (FAR). It is possible to change how close two feature vectors have to be and trade the FRR off against the FAR; it may be desirable to make the FAR very small (rarely give access to an impostor) at the expense of an increased FRR (legitimate users are refused more often) by demanding feature vectors be closer together.

A confusion matrix shows the difference ( $D$ ) between all pairs of a set of feature vectors. Figure 2.2 is the confusion matrix of an imaginary set of feature vectors describing three instances of three people. The difference between any vector with itself is 0 and normally the difference between different instances of the same person is less than the difference between feature vectors describing different people, i.e. there is a larger inter-class than intra class variance. The confusion matrix has reflective symmetry about the diagonal if  $D(A, B) = D(B, A)$ . The matrix is rescaled to 0-255 for a greyscale representation. The black pixels along the diagonal represent 0 difference between identical feature vectors, the dark diagonal 3 by 3 boxes show the similarity of instances of the same subject.

The central purpose of the system is to generate a feature vector that is unique to each person. For example, if performing beard recognition by extracting beard length in millimeters (mm) and hue from photographs of subjects, each pair is a feature vector and represents a point in a 2D feature space (beard space).

Subject	Instance 1	Instance 2	Instance 3	Instance 4	Instance 5
1	(7,16)	(10,15)	(12,17)	(13,15)	(13,16)
2	(107,22)	(103,23)	(99,20)	(105,21)	(101,20)
3	(20,16)	(23,17)	(19,17)	(26,15)	(21,18)

TABLE 2.1: Imaginary data from a beard recognition system.

Table 2.1 contains imaginary recognition feature vectors, each can be thought of as describing a point in a 2D feature space. If hue and length are good biometrics and can be captured accurately three clusters of points would emerge in the feature space representing the three subjects in Table 2.1. Should the clusters be sufficiently separated, distinguishing between these three subjects would be easy, the task of identifying an unknown subject is in fact the task of calculating which cluster the unknown point is closest to. Figure 2.2 is the confusion matrix generated from the points in Table 2.1, the self similarity of the subjects is apparent by the dark diagonal.

### 2.2.1 Different Biometrics

There are many types of biometric, measuring many different properties and behaviours. There are some common desirable properties of a biometric that contribute to its usefulness.

- Biometrics should be commonly available for measurement, a beard recognition system would fail when presented with the beardless or the slightly stubbly.
- Uniqueness. Biometrics that are unique to an individual give the ability to discriminate between subjects, we can not recognise people by the number of heads they have.
- Biometrics are less useful if they change rapidly over time, another pitfall for the newly shaven faced with beard recognition.
- A biometric that is easy to measure is good, toe print may be as good as a fingerprint but it requires more effort by the user or imagination by an engineer.
- If people object to a biometric it will limit its application. People may object to having their DNA recorded by their employer.
- It is of less use if a biometric can be easily faked, for example with a fake beard.

Different biometrics require different measurements and exhibit different strengths and weaknesses. Table 2.2 based upon [Lang and Leopold \(1997\)](#) compares several biometrics.

There are many other comparisons that can be drawn between biometrics and more than the three compared in Table 2.2, suffice to say no single biometric is perfect in

all situations and all have strengths and weaknesses, indeed the human mind with the array of biometrics it uses is far from infallible.

Physiological characteristics are in common use as biometrics, for example fingerprint or iris characterisation, but behavioural characteristics such as gait or language use are also used. There are many different biometrics and biometric applications.

Humans recognise each other by using facial characteristics. Face based biometric technology has to cope with variations in lighting, weather, expression and clothing but despite these complications, face recognition is possible ([Orozco-Alzate and Castellanos-Domínguez, 2007](#)). Face based biometrics have been used in such situations as tracking hooligans around football matches. Compared to other biometrics a person's face can be measured from long distances, however it is likely to change and relatively easy to disguise.

Voice is also used by humans to identify each other, however it is not expected that automated biometric systems will be able to uniquely identify people from large databases. Identifying a subject independent of the phrase spoken, or independent of the language spoken is difficult but has been performed ([Chaudhari and Ramaswamy, 2001](#)).

Fingerprint based identification is possibly the oldest biometric and was first used in China in the 14th century. Within a fingerprint there are characteristics referred to as minutiae, the spatial arrangement of which is unique and can be automatically compared ([Dass and Jain, 2006](#)). Fingerprint based identification is widespread and is used extensively in forensic applications, this has given it a criminal stigma that is now waning as it is increasingly used in diverse applications. It is planned that UK passports will contain a fingerprint biometric in addition to a facial biometric.

The texture of an iris is formed by chaotic processes during embryo development, the pattern is unique for each person and each eye. Iris based biometrics are extremely universal, unique and permanent, as well as being acceptable to most people. Its main disadvantages are that it can be disguised (but not easily copied), it requires some effort and co-operation by a user and can only be extracted at close range ([Daugman, 2004](#)).

Sections of DNA can be used to identify people ([Lake, 2007](#)), however it is currently restricted to forensic applications for several reasons; DNA processing is a complicated and relatively slow process, and capturing a DNA sample can be intrusive. There are also privacy concerns about compiling DNA databases as significant information about individuals may be gleaned from them, more so than any other biometric.

	Face	Fingerprint	DNA
Universality. Does everyone have one?	Good	Good	Good
Uniqueness. Is everyone's different?	Poor	Good	Good
Permanence. Does it change over time?	Medium	Good	Good
Collectability. Is it easy to measure?	Good	Medium	Poor
Range. Can it be measured from far away?	Medium	Poor	Poor
Acceptability. Do people approve of its use?	Good	Medium	Poor
Ease of Use. How easy is it to use?	Good	Medium	Poor

TABLE 2.2: A comparison of the performance of several biometrics. Adapted from [Lang and Leopold \(1997\)](#)

## 2.3 Gait Recognition

Gait recognition uses properties people exhibit whilst walking as a biometric. Several disciplines have historically had an interest in human gait. In psychological studies involving lights attached to body parts, humans have been able to identify the motion of gait and even other people ([Johansson, 1973](#); [Cutting and Kozlowski, 1977](#)). There are at least twenty distinct gait components such as thigh swing, shoulder twist etc. Most of these components have a large intra subject variance or can not be reliably extracted from video. Murray's work ([Murray, 1967](#)) however, supports the view that if all gait components are considered, human gait is unique. It follows that humans use some of these components of gait to identify each other, giving computers this ability is the aim of this research.

Potentially, gait has several advantages over other biometrics such as fingerprints and retinal images. Gait can be extracted from a distance with relatively low-resolution equipment and extracted from video automatically using computer vision techniques to subtract the background and find people in the image. As people have to walk from place to place, it is hard to avoid submitting a signature, not requiring user interaction also means users are not required to learn how to use the system. Measuring gait is acceptable to most people as it requires no physical contact, reveals no personal information and requires little additional action or effort by the user. Gait can even be used to covertly identify people without their knowledge and it is hard for a person to modify their gait without inconvenience and drawing attention.

However, as gait recognition is a relatively new area it is not currently known what the identity distinguishing capability of gait is. Also, gait is expected to change significantly

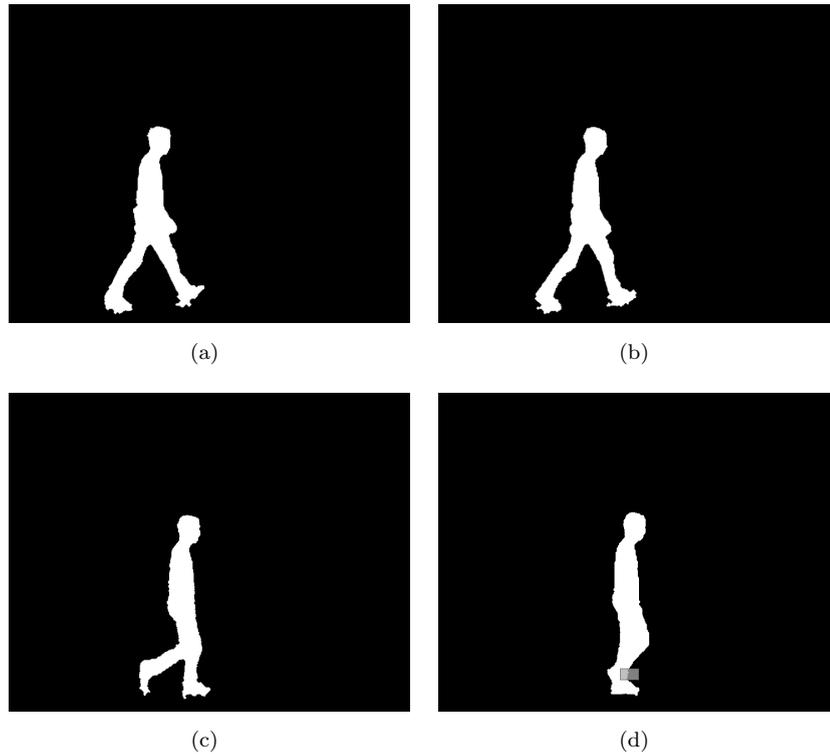


FIGURE 2.3: A heelstrike (a), double support (b), thighs occluding (c) and lower leg occlusion.

with injury and it is not known if other factors (e.g. mood, weather, terrain, clothing) can be compensated for. There are a number of challenges that need to be addressed before gait based biometrics are ready to be deployed in the real world.

There are several notable features of human gait: Figure 2.3(a) shows a heelstrike, this is when the heel or the forward extended leg first makes contact with the floor. Figure 2.3(b) is a frame taken during double support, when both feet are on the ground, during this time the legs are at their greatest deflection and the bounding box of the silhouette is at its greatest width. Figure 2.3(c) demonstrates the thighs becoming occluded, this is followed by the shins becoming occluded (Figure 2.3(d)). The self occlusion of the legs presents a challenge for those wishing to extract their position, it is not immediately obvious which leg the edge highlighted in Figure 2.3(d) belongs to.

Gait is a behavioural characteristic, a repeating periodic motion, as such the period of a subject's gait is often referred to, this is often measured using locally maximum width frames, or successive heelstrikes of the same leg.

A common experimental set-up for gait recognition has one camera with a straight track normal to the cameras view axis (see Figure 2.4). This set-up presents the maximal projected leg motion to the camera.

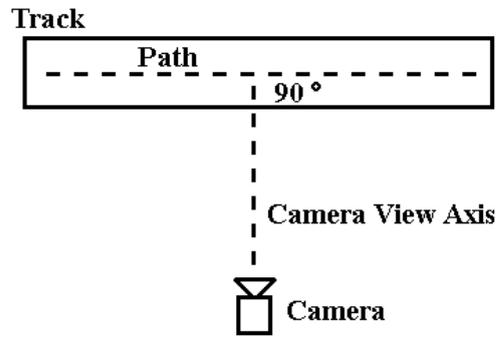


FIGURE 2.4: A common layout for a gait recognition experiment.

Gait recognition techniques can be broadly categorised as either model based or silhouette based (Nixon et al., 2005). Model based feature extraction methods use a representation of the human body to extract features from an image sequence. The use of a model helps extract features closely related to human gait such as the pendulum like motion of the thigh. Silhouette based techniques compute features of the silhouette shape such as its symmetry or width. This avoids the process of labelling the silhouette, but typically results in large feature sets with no specific correspondence between features and intuitive properties of gait.

### 2.3.1 Model Based Gait Recognition

Model based approaches to gait recognition attempt to extract information about the nature of a subject’s gait using the prior knowledge that they are a human who is walking, parameters describing the motion of the limbs (Cunado et al., 1999) or gait related constants (BenAbdelkader et al., 2002a). Silhouette based methods treat video of a subject walking more as that of an arbitrary shape to be classified.

The goal of gait recognition is to ascertain the identity of the person, not the pose or type of motion (which is usually known). This requires more accurate estimations of the motion and position of limbs. Consequently, model based gait recognition techniques use a subset of pose estimation methods that are tailored for accurate extraction of body properties (thigh motion, leg length etc.).

Static parameters such as cadence or periodicity of the walker and the stride length were used in BenAbdelkader et al. (2002a). Johnson and Bobick (2001) used height and stride length in addition to relative body parameters.

Intuitively, the motion of the thighs in the walking plane contains gait related information, the moving light displays people recognised in Johansson (1973), encoded this information. Sufficient information has been extracted from the movement of the thigh for recognition (Cunado et al., 1999). The approach was used in Cunado et al. (1999)

where an evidence gathering phase was performed with a Hough Transform to estimate the deflection of thighs through time ( $f(x)$ ) which was then encoded by a finite Fourier series described below.

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos n\omega t + b_n \sin n\omega t) \quad (2.1)$$

where  $a_n$  and  $b_n$  are the Fourier coefficients of  $f$ . The phase magnitude form of equation 2.1 is:

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} c_n \cos(n\omega t + \delta_n) \quad (2.2)$$

with

$$c_n = \sqrt{a_n^2 + b_n^2}, \quad n = 1, 2, \dots \quad (2.3)$$

and

$$\delta_n = \arctan\left(-\frac{b_n}{a_n}\right), \quad n = 1, 2, \dots \quad (2.4)$$

For a subject walking with a period of  $T$ ,  $\omega = \frac{2\pi}{T}$ . Either  $(a_n, b_n)$  or  $(c_n, \delta_n)$  pairs can be normalised by the magnitude of the first harmonic and used as features for gait recognition, only the first 5 harmonics  $n = 1 \dots 5$ , are required as it was found the higher frequencies are mostly noise (Cunado et al., 1999). Yam et al. (2001) extended this approach to the thighs and lower legs while running and walking and recorded recognition rates of 84 % upon 20 subjects using normal views from the Southampton large database. The Southampton large gait database (Shutler et al., 2002) has 2 views of over 100 subjects walking a linear track. One camera is situated normal to the track, the other approximately 25 degrees to normal. The database consists of sequences of silhouettes, background subtraction having been performed with the aid of chroma key extraction.

Rather than Hough Transform based evidence gathering a condensation based algorithm was used in Wang et al. (2004) to extract the joint angles. Others have used model based methods for analysing gait (Niyogi and Adelson, 1993; Tanawongsuwan and Bobick, 2001) as there are several advantages, most notably that properties of the subject's gait that are analysed and factors such as the appearance of clothing should have little impact on these measurements. By focussing upon a few measurements many other types of motion, such as that of the hips and the sideways motion of the legs, are ignored. Holistic silhouette based algorithms measure properties of the projected silhouette and in doing so the cumulative contribution of all visible aspects of gait.

### 2.3.2 Silhouette Based Gait Recognition

A human model is not required for gait recognition, most current gait recognition techniques are based upon established generic shape description algorithms (Veltkamp and Hagedoorn, 1999; Veltkamp, 2001). Selected significant frames were taken from the CMU Motion of Body database (Gross and Shi, 2001) and used as features (Collins et al., 2002), the database contained 25 subjects at 6 views evenly distributed around a treadmill.

Phillips et al. (2002b) used a silhouettes sequence itself as a feature, this was successfully tested on the Gait Challenge database (Phillips et al., 2002a). The Gait Challenge database contains two views of subjects walking an elliptical course.

Most methods, however, use a description of the silhouette or perform some other feature extraction to reduce the dimensionality of the feature vector and hopefully better encapsulate characteristics useful for gait recognition.

Moment based silhouette descriptions (Shutler and Nixon, 2001; Little and Boyd, 1998) have been used successfully for gait recognition as have features generated from simple measures of sub sections of the silhouette (Foster et al., 2001; Lee, 2001).

The changing nature of the silhouette can be described by its projection into 1D functions. In Kale et al. (2002, 2003b) silhouettes are characterised by their changing width, frieze patterns (Liu et al., 2001) are created by summing the columns or rows and collating through time. Frieze pattern analysis was performed upon the University of Maryland Human ID Database (HID-UMD) (Kale et al., 2003b). HID-UMD contains walking sequences of 25-50 people at 2 orientations, normal and frontal views.

Gait is periodic and symmetric in nature, BenAbdelkader et al. (2002b) uses temporal symmetry and (Hayfron-Acquah et al., 2002) combines temporal with spatial symmetry, this was tested upon the Southampton large database.

## 2.4 Orientation Independence

As gait is a relatively new area of interest, most algorithms designed to measure gait and perform recognition operate in highly constrained environments. Most control the path the subjects travel along, some control the lighting and the background and a few control the clothing and appearance of the subjects. This work has shown that automatic gait recognition is possible with a variety of techniques, however for practical implementations out of the laboratory these constraints must be relaxed; useful systems will require a greater degree of independence with respect to subject trajectory, appearance and illumination.

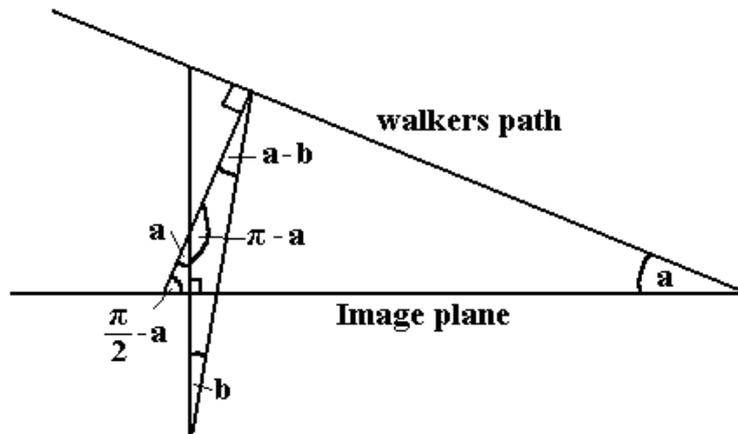


FIGURE 2.5: Given the track angle  $a$  to the image plane and a ray from the subject at angle  $b$  to the camera's view axis the orientation of the subject is  $a - b$

Most gait recognition algorithms have been developed and tested with data captured from cameras normal to the direction the subject is walking. This view presents the largest possible visible motion of the legs. Typically, measurements extracted change when taken from non-normal orientations. In a real system, subjects can be expected to appear at a wide range of orientations, even a person walking along a straight path normal to the camera's view axis will be changing orientation with respect to the camera and will only briefly be approximately normal <sup>1</sup>.

With knowledge of the orientation of the path that the walker is travelling with respect to the image plane, the orientation of the subject at each point along it can be calculated. Figure 2.5 shows the relationship between the angles involved. With a path at angle  $a$  to the image plane and a ray from the centre of projection to the subject at angle  $b$  to the optical axis, the orientation ( $o$ ) of the subject with respect to the camera is  $a - b$ .

Assuming  $a$  is known we can calculate  $b$  from the subject's position in the video frame  $(u, v)$  and the perspective projection matrix Equation 2.5;

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{Z} \begin{bmatrix} a_u & 0 & u_c & 0 \\ 0 & a_v & v_c & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (2.5)$$

where  $a_u = \frac{f}{\text{pixel width}}$  and  $a_v = \frac{f}{\text{pixel height}}$ . There are five parameters describing the camera, the focal length ( $f$ ), the pixel width and height,  $u_c$  and  $v_c$  co-ordinate of the

<sup>1</sup>There is weak evidence for there being an effect due to orientation, the average image algorithm (Veres et al., 2004) recorded lower recognition rates when comparing sequences of subjects walking right against those walking left (Carter, 2008) and attribute this to the shapes of the walkers deforming in different ways.

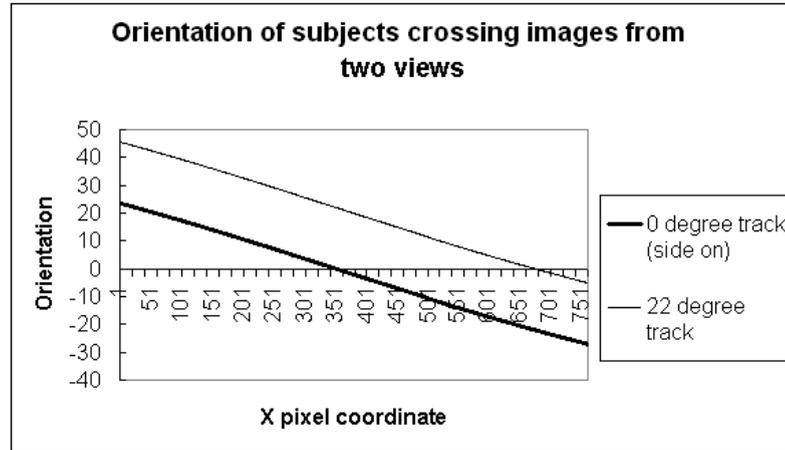


FIGURE 2.6: The orientation of a subject as they cross the camera's field of view

optical centre. Of the  $X, Y, Z$  co-ordinate system  $X, Y$  are parallel to  $(u, v)$  with the origin at the intersection of the  $(u, v)$  image plane and the optical axis. From Equation 2.5 a subject's orientation in the  $X-Z$  plane is

$$u = a_u X/Z + u_c \quad (2.6)$$

which can be rearranged to

$$\frac{u - u_c}{a_u} = X/Z \quad (2.7)$$

A subject at  $x, y, z$  is at angle  $b$  in Figure 2.5, where  $b = x/z$  so the equation for the orientation of a subject can be written in terms of the camera calibration parameters  $u_c, a_u$ , the horizontal pixel coordinate ( $u$ ), and the angle of the track with respect to the  $u, v$  image plane  $a$ .

$$\text{orientation} = a - \arctan \frac{u - u_c}{a_u} \quad (2.8)$$

Using the camera calibration parameters for the two cameras used to film the Southampton large database Figure 2.6 was created, it shows the orientation of a subject change by approximately 50 degrees. Figure 2.7 shows the cosine of these angles in Figure 2.6. The primary motion of the legs is scaled by this amount, however, due to each leg swinging in a different plane, correctly rescaling the legs is not simple.

Very few of the existing gait recognition techniques are designed for use over a range of orientations, normally subjects walk in a path orthogonal to the cameras view axis.

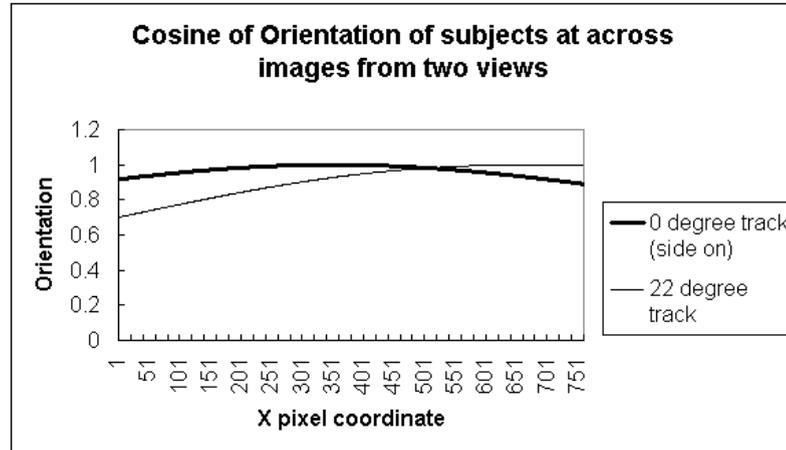


FIGURE 2.7: The horizontal rescaling of a subject as they cross the camera's field of view.

During recognition most have the prerequisite that there is a gallery at the same orientation as the probe sequence. There are several approaches to introducing orientation independence into a gait recognition algorithm.

- Assume the subject will at some point present a normal view to the camera. Ignoring the problem may be the best strategy in some situations. If people can be forced to walk designated paths e.g. down a corridor or through a tunnel then you can guarantee an ideal view. Nevertheless, all the non-normal view data will be discarded.
- Synthesise a normal view. By modelling people as planar surfaces it is possible to synthesise a normal view, this relies on people being approximately planar. With multiple cameras it may be possible to build a sufficiently accurate 3D representation that can be used to generate increasingly accurate silhouettes.
- Record a gallery containing features recorded near all possible orientations. A gallery can not be recorded for the infinite number of orientations. If a gallery is present that is sufficiently close to the orientation of the probe the effect of orientation may be marginal. Capturing galleries 'sufficiently close' to all possible orientations may not be practical.
- Transform features generated from a non-normal orientation to those from a normal orientation. This will be dependant on the nature of the features.
- Generate features that are independent of orientation. It is very hard to extract features that are independent of orientation, people appear significantly different between front and side views.

A combination of the above approaches may be used, for example, a technique may generate parameters that change by a sufficiently small amount over a range of  $5^\circ$ ,

so several galleries may be recorded over a range of  $\pm 30^\circ$  from normal at five degree intervals. Cameras can then be positioned and the subject's path constrained so a view in the range  $\pm 30^\circ$  is always visible. This is dependant upon a technique that is orientation independent over  $5^\circ$ , the ability to capture the multiple views for the gallery, accurate determination of the subject's orientation and a scenario where probe cameras can be placed and the path constrained. A technique that is independent of orientation over a larger range of views requires fewer galleries to be captured, and more flexibility in the scene set up.

Some research has been conducted into increasing orientation independence of gait algorithms. Body parameters such as height and stride length have been translated to a normal view using motion captured data from a reference subject ([Johnson and Bobick, 2001](#)), however, this is difficult for other features of gait. Another possible approach is to transform the silhouette to a normal orientation before measurements are taken, [Chowdhury et al. \(2003\)](#) models people as planar surfaces and can normalise silhouettes until this approximation fails. [Matusik et al. \(2000\)](#) details how to generate visual hulls from multiple cameras, [Lee \(2001\)](#) uses this to generate normal silhouettes from a visual hull generated from multiple cameras. [Spencer and Carter \(2002\)](#) and [Spencer and Carter \(2005\)](#) follows on from [Carter and Nixon \(2000b\)](#), normalising measured thigh angles for model based recognition. Calculating the normal thigh angles requires accurate measurements of key body points through the video sequence, this was achieved with markers attached to the subjects. [Laxmi et al. \(2002\)](#) demonstrated that by using the 3D motion capture data ([Geisheimer, 2001](#)) to generate features, orientation independent recognition was possible, however this required a motion capture studio.

Using static properties of the silhouettes ignores a wealth of information about the motion of the subject. Trading useful features for orientation independence is not a desirable solution as gait does not currently have the distinguishing ability of other biometrics. Approximating people with planar surfaces only works whilst people are approximately planar. Accurate Visual hulls require multiple cameras, this is not always viable and further hindered by the dynamic complex self occlusion of a subject in motion.

## 2.5 Contribution.

This thesis shows the importance of subject orientation with respect to gait recognition. We show that any silhouette based algorithm will be effected by subject orientation, and that there are significant changes in orientation even in tightly controlled experimental setups.

Three silhouette based gait recognition algorithms are examined and found to fail when comparing subjects differing by 20 degrees. The complex changes upon a silhouette

caused by changing orientation are detailed and a human model is required to interpret silhouettes between multiple orientations.

A novel model based recognition algorithm is developed and shown to have an increased tolerance to changing orientation than the silhouette based methods.

## Chapter 3

# Baseline Algorithm at Multiple Orientations

In this chapter, the performance of a baseline gait recognition algorithm is examined with subjects at a range of orientations. In these experiments the Human ID Baseline algorithm (Phillips et al., 2002b) was used, this is the accepted baseline comparison algorithm for gait recognition. This algorithm uses raw silhouettes with very little processing to generate feature vectors, its simplicity implies any effects are due to the silhouettes rather than the nature of the algorithm. This set of experiments gives an insight into what happens to a simple algorithm and can be used as a baseline for comparison with other algorithms.

None of the datasets described in Section 2.4 or Section 2.3 contain subjects walking at a wide range of orientations. To explore the effect of orientation artificial silhouettes were generated at 19 different orientations using an OpenGL manikin animated using motion capture data.

### 3.1 Southampton GTRI Artificial Walker

The Southampton GTRI (Georgia Tech Research Institute) Artificial Walker (SGAW) was created to enable fine grained exploration of orientation effects upon gait recognition. The SGAW is an OpenGL articulated person animated with 3D motion capture data (Geisheimer, 2001) of real subjects walking. Using the walker, silhouettes of a subject at any orientation can be generated for experimentation.

The motion capture data consists of x,y,z coordinates of markers attached to the body, Figure 3.1(a) shows the points known. OpenGL with Python was used to draw 3D cylinders between these points, gaps between body parts were filled with spheres at joints. Figure 3.1(a) shows the resulting layout of spheres (circles) and cylinders (rectangles).

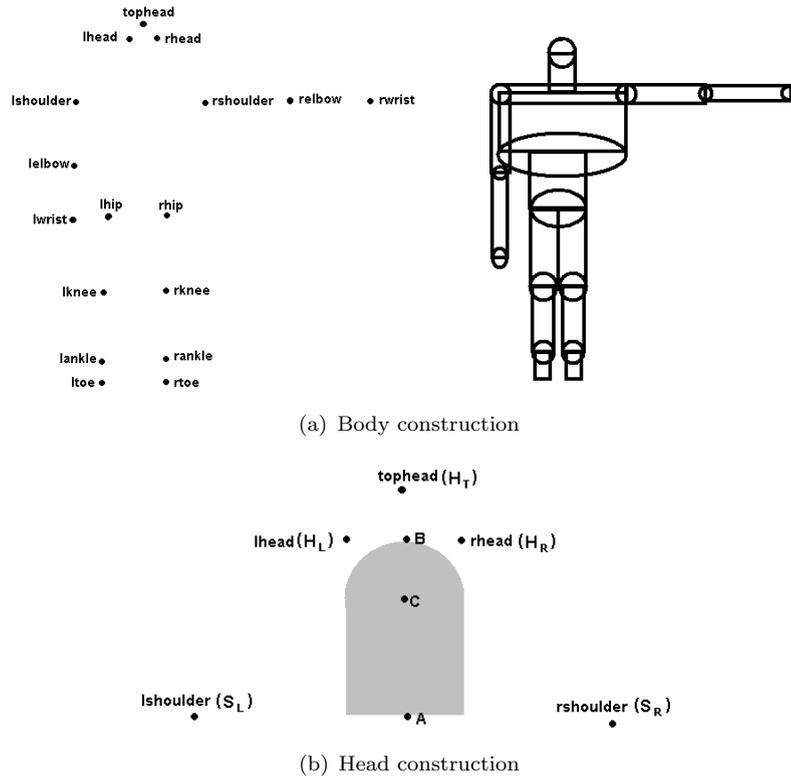


FIGURE 3.1: The motion capture points and the construction of the 3D model.

Body part	Width
Thigh	1.0
Lower Leg	0.8
Foot	0.5
Upper arm	0.6
Lower arm	0.6

TABLE 3.1: The heuristics controlling the width of body parts.

Half the hip width is used as the thigh width, Table 3.1 specifies the widths of the other body parts as a fraction of the thigh width. Although crude, using this heuristic created realistic looking silhouettes.

The shape of the body is unknown so two cylinders are used, one between the shoulders and one between the hips with a sphere at the joint to smooth the transition.

The silhouettes generated were as noise free as possible, also body parts that were not considered relevant could be removed (Figure 3.2(b)). Due to uncertainties about the exact placement of the markers during motion capture, some properties of the SGAW such as thigh length (Figure 3.2(a)) appear inaccurate. This is not a large concern as the nature of gait is encapsulated in the video even if the silhouettes generated are marginally anatomically unrealistic.

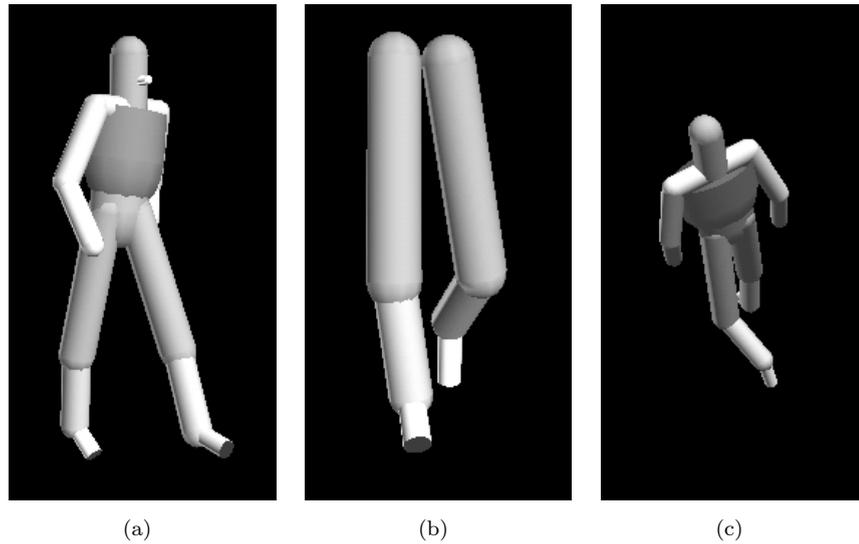


FIGURE 3.2: Some examples of the data that can be produced using the 3D model.

Using SGAW generated silhouettes the hips can be fixed in the centre of the image, as a result the subject is at a constant distance from the camera so rescaling silhouettes before any processing is not necessary, also silhouettes need not be centred or aligned. Fixing the hips and making the manikin walk on the spot removes the effect of a changing orientation due to subjects passing the camera as shown by Figure 2.6 in Section 2.4.

The period of a subject's gait can be estimated directly from the motion capture data by calculating the time between successive periodic instantaneous events that feature in normal gait. A good feature would be a thigh passing the vertical as it swings forward. However, a preferable method is to calculate the period from the generated silhouettes, then should these experiments be repeated with live captured video the same period estimating algorithm can be applied.

### 3.2 Calculating the Period

The period of a walker's gait is the length of time before their motion repeats. To measure the period of a walker's gait some distinctive point in their stride is noted and the mean time between successive occurrences calculated. For example, the mean time between successive heelstrikes of the left foot. Knowing the period of a subject's gait allows the analysis of single stride or an integer multiple thereof. Statistics pertaining to sequences of frames that are not an integer multiple of the period may be different depending on where in the stride the sequence begins.

To calculate the period of a set of silhouettes, the following procedure was implemented. Bounding box widths ( $W$ ) were extracted (of  $N$  frames  $W = w_0, \dots, w_N$ ) from silhouette

sequences, a function ( $f$ ) is then fitted using a standard library curve fitting function; Levenberg-Marquardt optimisation, where;

$$f(x) = |\cos(2 * \pi * x/a) + b) * c| + d \quad (3.1)$$

$a$  is the period of the sequence,  $b$  is the phase of the sequence,  $c$  is the amplitude of the signal and  $d$  is a value akin to the DC offset, the minimal width of the silhouette.  $|\cos|$  is used to model the width, this is based upon the observation that the width will largely be determined by the swinging of the legs, as the legs cross whilst travelling at their maximum speed, a sharp change in the rate of change of the width is expected.  $|\cos|$  better models this than  $\cos$ . To initialise these values frames  $w_i$  are found with minimum widths over a range of  $w_{i-5}, \dots, w_{i+5}$  frames. Figure 3.3 shows the widths from a silhouette sequence. Experience has shown that minimal widths are more robust to image noise and more temporally consistent than maximal values. The change in width near the minimas is not as sharp as that of  $|\cos|$ , this may be due to the sampling rate of the camera and the non sinusoidal nature of the leg motion. People don't walk with a period of less than 20 frames (at 25 fps), consequently minima are expected to occur with double the frequency of their gait so the minimal value in a range of 10 frames is appropriate.

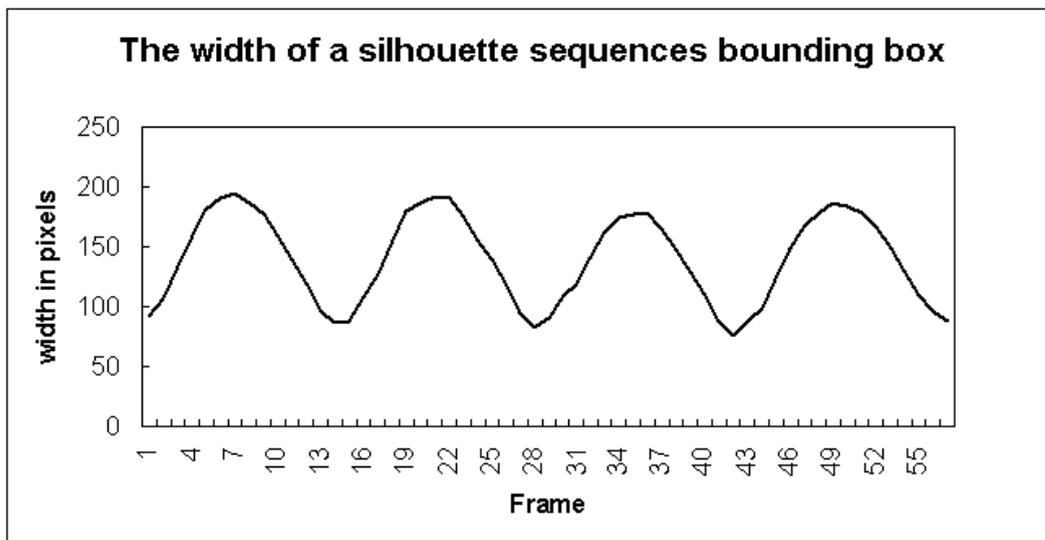


FIGURE 3.3: Measured widths of bounding boxes of a silhouette sequence.

Using the indices of the frames with locally minimal widths ( $V = i_1, \dots, i_M$ ) initial values  $a_0, b_0, c_0, d_0$  are estimated for the optimisation:

$$a_0 = 2 * \left( \sum_{j=2}^M V_j - V_{j-1} \right) / (M - 1) \quad (3.2)$$

$$b_0 = 2 * \pi * \frac{V_0}{a_0} + \frac{\pi}{2} \quad (3.3)$$

$$d_0 = \left( \sum_{j=1}^M W_{V_j} \right) / M \quad (3.4)$$

$$c_0 = 2 * \left( \sum_{i=1}^N W_i / N - d_0 \right) \quad (3.5)$$

$a_0$  is the mean period calculated from the frame numbers of the minima,  $b_0$  is calculated using the initial estimate of the period  $a_0$  and the frame number of the first minima, this does make the initialisation sensitive to the correct estimation of the first minima however this was not a problem in practice.  $d_0$  is initialised as the mean width at minima and  $c_0$  is initialised as double mean difference from  $d_0$ .

Figure 3.4 shows the function and the function once the parameters have been optimised. Equation 3.1 appears to be a good model of the changing width of a silhouette. The error that remains is due to the non sinusoidal motion of a walker's limbs, the width of limbs themselves and the effect of a changing view upon the projected width. Using this technique to estimate phase and period from real data was successful. 1600 sequences were processed from the Southampton large database, inspecting a sample of the width traces combined with the optimised function (as in Figure 3.4) found the algorithm robust and accurate. Also the period was compared to a manually derived estimate based upon marking successive heel strikes, this agreed with the calculated period.

There is room for improving the model in Equation 3.1 to more accurately describe the changing width of the subject, though it is unlikely this would yield significant improvement in the accuracy of the phase and period estimation.

### 3.3 Original Algorithm

Using the SGAW to generate silhouettes, the baseline algorithm can be applied to compare silhouette sequences to determine whether they are of the same subject. The baseline algorithm from (Phillips et al., 2002b) is described below.

Given a probe sequence ( $P$ ) and a  $m$  long gallery sequence ( $G$ ), first partition  $P$  into disjoint sub-sequences ( $P_k$ ) containing  $n$  contiguous frames ( $P_k(1), \dots, P_k(n)$ ) where  $n$  is approximately one stride in length,  $n \approx p * fr$ ,  $p$  being an estimate of the period of a subjects gait and  $fr$  the frame rate of the camera. Each sub-sequence ( $P_k$ ) is correlated with the gallery sequence at several temporal offsets ( $o$ ):

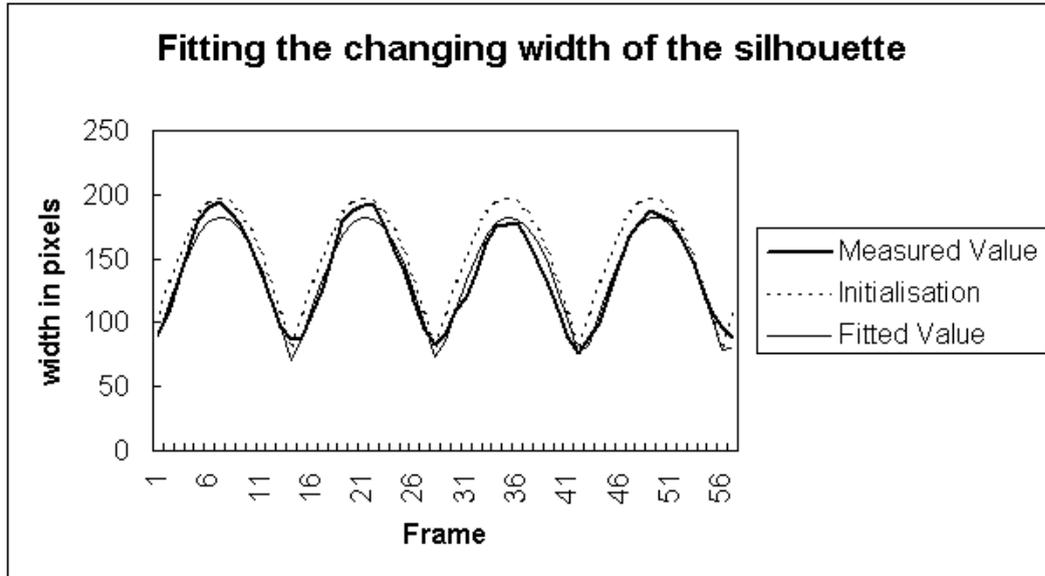


FIGURE 3.4: Measured widths, an initialised curve and the final fitted curve.

$$Corr(P_k, G, o) = \sum_{i=1}^n Dist(P_k(i), G(i+o)) \quad (3.6)$$

where  $Dist$  is the distance between two silhouette frames.  $Dist$  is the ratio of the number of frames in their intersection to the number of frames in their union:

$$Dist(P(i), G(j)) = \frac{Num(P(i) \cap G(j))}{Num(P(i) \cup G(j))} \quad (3.7)$$

where  $\cap$  is the intersection of the silhouettes and  $\cup$  is the union, unintuitively  $Dist$  increases as the feature vectors are closer. The offset ( $o$ ) is used to compare probe sub-sequence ( $P_k$ ) with all parts of the gallery sequence ( $G$ ), the maximum value of  $Corr(P_k, G, o)$  is chosen to be the correlation of the sub-sequence with the gallery. The similarity of  $P$  and  $G$  is chosen to be the median correlation of all sub-sequences.

$$Sim(P, G) = Median_k \left( \max_{o=0}^{m-n} [Corr(P_k, G, o)] \right) \quad (3.8)$$

### 3.4 Modified Version

A modified version of the Phillips et al. (2002b) algorithm is used to make it less computationally demanding. The phase of the subject's gait was estimated with the period in Section 3.2 using this, a sub sequence of frames was selected that was one period

in length and started at the same point in the gait cycle. Having phase aligned sequences makes comparisons at many offsets unnecessary. Additionally, rather than split the probe into several sub-sequences a single period is used, the gallery also contains a single period of silhouettes.

Unfortunately the length of the probe ( $p$ ) and the length of gallery ( $g$ ) may be different, to overcome this  $Corr$  is calculated over  $p$  frames. This may mean that some gallery silhouettes are ignored, or if  $o > g - p$  the gallery will be lacking silhouettes. In the latter case frames are wrapped around from the end of the gallery by indexing with  $o + i \bmod g$ . This will create discontinuities in the silhouette sequences as the first and last silhouette will be different due to the environment and perspective effects even though they are similar points of the stride.

Equation 3.6 modified:

$$Corr(P, G, o) = \sum_{i=1}^p Dist(P(i), G(o + i \bmod g)) \quad (3.9)$$

The distance between two silhouette frames  $Dist$  remains unchanged from Equation 3.7. The number of offsets tried in the similarity equation Equation 3.7 is reduced from  $0 \dots m - n$  to a much smaller set:

$$Sim(P, G) = \max_{o \in O} (Corr(P, G, o)) \quad (3.10)$$

There is no need to use the median as there is only one stride in the probe and gallery. Initially -1, 0 and 1 are used as offsets ( $O = [-1, 0, 1]$ ), if using  $o = 0$  gives the highest (best) distance of the three offsets, the  $Corr$  with  $o = 0$  returned. If 1 or -1 gives the maximum  $Corr$ , the set  $O$  is extended with the next integer offset in that direction while each new  $o$  gives a better  $Corr$ .  $O$  may finish as  $[-1, 0, 1, 2, 3, 4]$  with  $o = 3$  giving the maximum  $Corr$ . This is a very simple 1D optimization, that is essentially choosing the initial frame given that the estimates of phase and period have sub-frame accuracy. A more complex 1D optimization algorithm could be used such as the golden section or Brents method but as the optimal  $o$  is expected to be 0 or close to 0. This follows from the fact that the sequences are all aligned using the pre-calculated phase and period of subjects does not differ greatly.

### 3.5 Experimental Setup.

The modified algorithm was used with silhouettes generated with the SGAW described in Section 3.1. Silhouettes of 4 instances of 19 subjects walking one complete stride at 19 orientations (0, 5, 10, ..., 85 and 90 degrees to the normal view, 90 degrees being

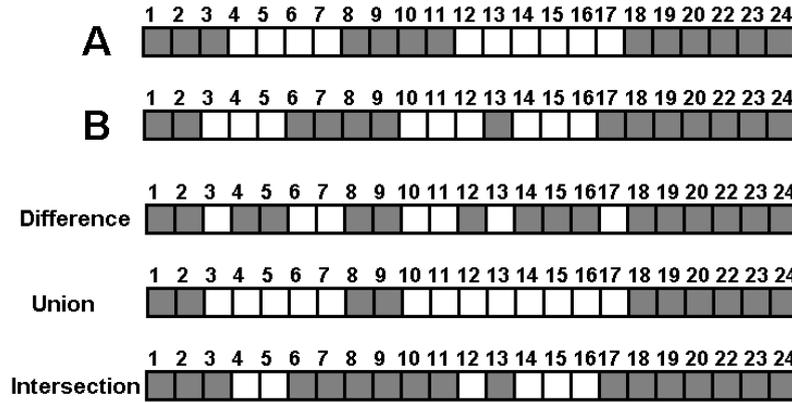


FIGURE 3.5: Encoding a flat binary image.

frontal) The size of the dataset is considerable, using approximately 50 frames of 150 by 300 images of 4 instances of 19 people at 19 orientations at 1 byte per pixel uses 3.2 Gigabytes, comparing two sequences typically involves comparing a minimum of  $3 * 50 * 150 * 300 = 6750000$ , 8 bit integers. Packing the integers with 8 pixels per byte would reduce the size to 0.4 Gig but add complexity to comparison functions. Instead a run length encoding based scheme was used to losslessly reduce the size of the data and speed comparisons.

Each image was reshaped into a 1D array, the location of each pixel different to the one before it was recorded. For example, array A in Figure 3.5 would be encoded [4, 8, 12, 18] and B, [3, 6, 10, 13, 14, 17]. The image can be flattened along either axis although with human silhouettes flattened along the vertical axis (concatenating columns) gave the best compression, having the smallest number of contiguous blocks.

The number of pixels ( $Num$ ) in an encoded image  $I$  is:

$$num(I) = \sum_{i=0,2,\dots} I_{i+1} - I_i \quad (3.11)$$

and the difference (XOR) between two encoded images  $I^1$  and  $I^2$  is:

$$dif(I^1, I^2) = sort(e^1 \cup e^2) \quad (3.12)$$

where  $sort$  is any sorting algorithm. From Figure 3.5:

$$dif(A, B) = sort([4, 8, 12, 18] \cup [3, 6, 10, 13, 14, 17]) \quad (3.13)$$

$$dif(A, B) = sort([4, 8, 12, 18, 3, 6, 10, 13, 14, 17]) \quad (3.14)$$

$$dif(A, B) = [3, 4, 6, 8, 10, 12, 13, 14, 17, 18] \quad (3.15)$$

It may be possible to create a very efficient *sort* as  $I^1$  and  $I^2$  are known to be in ascending order. If both encoded images start or end a block at the same point there will be repeated indices in the resulting encoding, these can safely be removed.

To generate the intersection of two encoded images A and B the following algorithm is used:

```
intersection = new list
# initialising the states
Ais1 = false
Bis1 = false
IntersectionIs1 = false

#go through A and B updating the states and creating the intersection
while there are more points in A and B:
    pop the next point from A or B
    update Ais1 or Bis1
    if Ais1 and Bis1
        add this point to the intersection list and set IntersectionIs1 = true
    else if IntersectionIs1:
        add this point to the intersection list and set IntersectionIs1 = false

return intersection
```

The union is very similar to the intersection and can be merged with the intersection algorithm. If the number of pixels in the intersection or union is desired this too can be carried out during the same loop. This run length based representation gives much smaller storage and faster computation of *Dist* described by Equation 3.7.

## 3.6 Results

Four experiments were performed, the first two examined the effect of orientation upon the similarity of silhouettes. It is expected that instances of the same subject will have similar silhouettes and that silhouettes of different subjects will differ more. The first two experiments look at how the orientation of the subjects effects the inter and intra subject silhouette similarity.

Two were performed to examine the performance of the algorithm at different orientations, two recognition experiments were then carried out. Four instances of nineteen people at a range of orientations (0, 5, 10, 15...90 degrees to the normal view) were encoded and used.

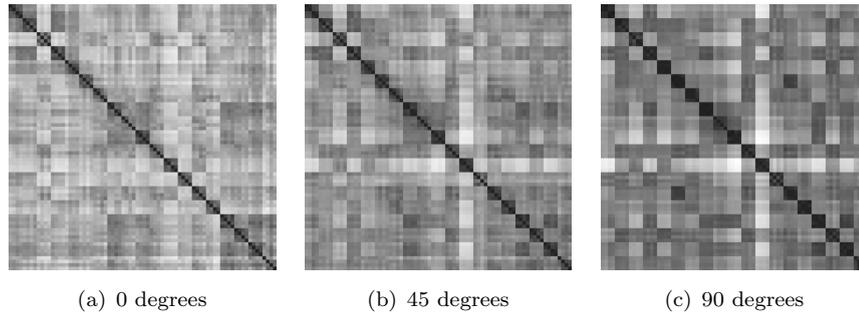


FIGURE 3.6: The confusion matrix of all runs of all subjects at 0,45 and 90 degrees. 4 by 4 blocks represent a comparison of 4 instances of one subject with 4 instances of another. Dark 4 by 4 boxes along the diagonal indicate there is little difference between instances of the same subject and a large difference between subjects.

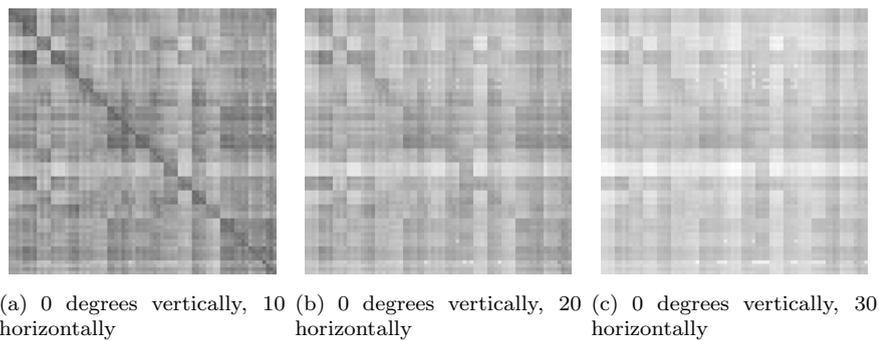


FIGURE 3.7: The confusion matrix of all runs of all subjects at 0 degrees vs other orientations.

The first experiment consisted of calculating the similarity of all instances of all subjects at the same orientation. The results are presented in confusion matrices with *Sim* rescaled to the range  $0 \dots 255$  (Figure 3.6 and Section A.1). The dark diagonal squares of 4 by 4 pixel groups indicates the small distance between instances of the same subject, the larger difference between different subjects is shown by the comparative lightness of off diagonal comparisons. The confusion matrices show that when using a probe and gallery at the same orientation, there is a significantly larger inter subject variance than intra subject variance and that if the orientation of all the compared silhouettes is the same, it matters little what the actual orientation is.

The second experiment shows confusion matrices generated from silhouettes of subjects at  $0^\circ$  compares with subjects at  $5, 10, 15 \dots 90$  degrees (90 being in front of the subject), the confusion matrices are presented in Figure 3.7 and Section A.2. As the difference in orientation increases, the dark diagonal visible in Figure 3.6 lightens as the intra subject silhouette similarity becomes similar to the inter subject similarity. In other words, silhouettes of a subject captured at  $0^\circ$  become increasingly similar to all other subjects as the orientation of the silhouettes they are compared to increases.

Probe Instance	Probe and Gallery Orientation									
	0	5	10	15	20	25	30	35	40	45
1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
3	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Mean	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Probe Instance	Probe and Gallery Orientation								
	50	55	60	65	70	75	80	85	90
1	100.0	100.0	100.0	94.7	94.7	94.7	94.7	94.7	94.7
2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
3	100.0	100.0	100.0	100.0	94.7	94.7	94.7	94.7	89.5
4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Mean	100.0	100.0	100.0	98.7	97.4	97.4	97.4	97.4	96.1

TABLE 3.2: Recognition rates, when the probe and gallery are at the same orientation.

The third and fourth experiments were recognition experiments. For a recognition experiment a sequence of silhouettes referred to as the probe has an unknown identity, the gallery is a set of reference silhouette sequences with known subject identities. The algorithm attempts to correctly ascertain the identity of the probe by calculating which sequences of the gallery, and consequently subject identities, the probe is most similar to. The gallery does not contain the probe but does contain other instances of that subject. The recognition succeeds if the calculated identity is that of the probe, it fails otherwise. The data being used is that of 4 instances of 19 subjects, each instance of each subject is used in turn as the probe, the gallery consists of the other 3 instances of that subject and 3 instances of the other 18 subjects. The recognition rate is the percentage of probe that is successfully identified.

The third experiment simulates the situation where the orientation of the probe is known and there is a gallery of silhouettes at every orientation available, hence the orientation of the silhouettes in the gallery is the same as that of the probe. Recognition is performed using K-NN with  $K = 3$ .

Table 3.2 contains the recognition rates observed with probes and galleries at each orientation. The table also contains the mean recognition rate at each orientation, this is plotted in Figure 3.8. Recognition appears relatively easy, with only a small error at orientations close to a frontal view. This error may be due to the incorporation of motion in the Y-Z plane that is unobservable at the normal view.

The fourth experiment is more comprehensive, galleries recorded at all orientations (0,5,...,90 degrees) were used in turn to identify probes at all orientations. For each combination of probe and gallery orientation each of the 4 instances were again used as probe and the corresponding instance removed from the gallery. When the probe orientation and gallery orientation is identical it is in effect the same experiment as

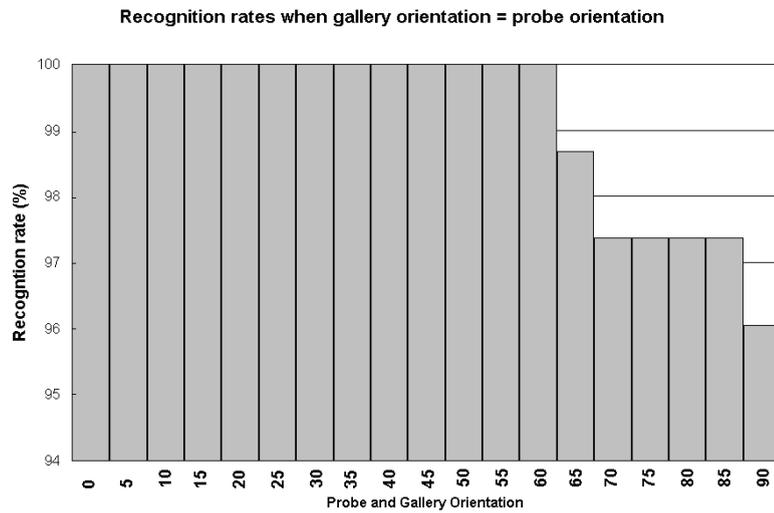
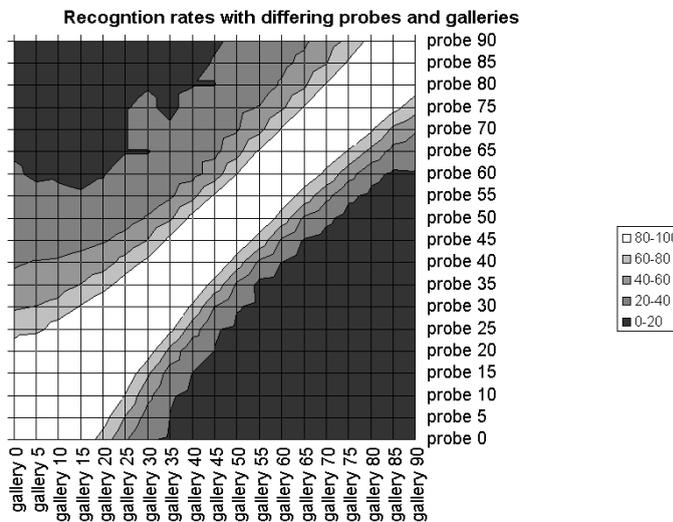


FIGURE 3.8: The recognition rate at each orientation, where, orientation of the probe = orientation of the gallery



(a)

FIGURE 3.9: The recognition rate with galleries and probes at different orientations.

above, as the difference in orientation increases, recognition rates are expected to drop. The second experiment showed that for small differences in orientation the inter subject similarity remained much smaller than the intra subject similarity, but this did not hold for large differences in orientation.

Figure 3.9 and Table 3.3 describe the recognition rates with probes and galleries at various orientations. Recognition rates fall quickly as the angle between probe and gallery increases, a difference of 20 degrees results in a drop of 20%. Figure 3.9 shows that there is a greater tolerance of a change in orientation near the normal views, at orientations greater than 45 degrees a difference in orientation of 10-15 degrees results in a drop of 20%. This is most likely due to the fact that the apparent motion of the

Probe Orientation	Gallery Orientation									
	90	85	80	75	70	65	60	55	50	45
90	96.1	94.7	86.8	64.5	50.0	36.8	31.6	26.3	22.4	18.4
85	94.7	97.4	94.7	82.9	59.2	44.7	35.5	28.9	25.0	21.1
80	89.5	96.1	97.4	94.7	81.6	56.6	42.1	28.9	27.6	19.7
75	68.4	86.8	94.7	97.4	96.1	81.6	56.6	40.8	35.5	26.3
70	42.1	55.3	82.9	92.1	97.4	97.4	81.6	53.9	38.2	32.9
65	27.6	32.9	53.9	78.9	90.8	98.7	98.7	82.9	51.3	35.5
60	18.4	17.1	25.0	47.4	76.3	92.1	100.0	98.7	80.3	48.7
55	15.8	11.8	15.8	23.7	44.7	72.4	93.4	100.0	97.4	82.9
50	7.9	6.6	9.2	13.2	23.7	36.8	72.4	94.7	100.0	98.7
45	5.3	5.3	5.3	9.2	13.2	18.4	36.8	72.4	94.7	100.0
40	5.3	5.3	5.3	5.3	7.9	11.8	19.7	35.5	71.1	97.4
35	5.3	5.3	5.3	5.3	5.3	9.2	11.8	14.5	36.8	72.4
30	5.3	5.3	5.3	5.3	5.3	6.6	10.5	15.8	21.1	39.5
25	5.3	5.3	5.3	5.3	5.3	5.3	7.9	14.5	17.1	21.1
20	5.3	5.3	5.3	5.3	5.3	5.3	7.9	11.8	13.2	19.7
15	5.3	5.3	7.9	6.6	7.9	7.9	7.9	10.5	14.5	14.5
10	7.9	7.9	5.3	5.3	5.3	5.3	6.6	10.5	11.8	13.2
5	7.9	6.6	5.3	5.3	5.3	5.3	5.3	5.3	6.6	11.8
0	7.9	7.9	5.3	5.3	5.3	5.3	5.3	5.3	6.6	10.5

Probe Orientation	Gallery Orientation									
	40	35	30	25	20	15	10	5	0	
90	14.5	17.1	14.5	11.8	11.8	13.2	10.5	13.2	15.8	
85	15.8	15.8	15.8	17.1	15.8	15.8	9.2	11.8	13.2	
80	19.7	18.4	19.7	18.4	15.8	15.8	13.2	11.8	11.8	
75	22.4	18.4	21.1	19.7	17.1	14.5	13.2	11.8	11.8	
70	25.0	21.1	22.4	19.7	18.4	14.5	14.5	14.5	14.5	
65	32.9	22.4	19.7	19.7	17.1	15.8	15.8	15.8	18.4	
60	32.9	28.9	22.4	22.4	19.7	17.1	18.4	17.1	22.4	
55	53.9	36.8	25.0	22.4	21.1	21.1	22.4	25.0	22.4	
50	86.8	56.6	42.1	31.6	27.6	23.7	28.9	26.3	26.3	
45	98.7	85.5	60.5	46.1	38.2	34.2	31.6	30.3	32.9	
40	100.0	98.7	85.5	67.1	51.3	46.1	42.1	40.8	38.2	
35	97.4	100.0	98.7	90.8	72.4	60.5	51.3	48.7	44.7	
30	76.3	100.0	100.0	98.7	94.7	81.6	65.8	60.5	57.9	
25	48.7	85.5	100.0	100.0	100.0	96.1	89.5	75.0	73.7	
20	25.0	53.9	90.8	100.0	100.0	100.0	98.7	97.4	88.2	
15	19.7	30.3	63.2	96.1	100.0	100.0	100.0	100.0	98.7	
10	17.1	21.1	44.7	78.9	98.7	100.0	100.0	100.0	100.0	
5	15.8	19.7	31.6	57.9	88.2	100.0	100.0	100.0	100.0	
0	13.2	18.4	21.1	42.1	72.4	94.7	100.0	100.0	100.0	

TABLE 3.3: Recognition rates for when the probe and gallery are at different orientations.

legs changes approximately with the cosine of the angle measured from the side view. A similar effect upon torso may explain the marginally wider range near 90 degrees. Also of note is the non-symmetry of the recognition across the diagonal reflecting the non-symmetry of the comparison algorithm. The modified version of the algorithm may be introducing this artefact when short gallery sequences are 'wrapped' for comparison with longer sequences.

### 3.7 Conclusions

It was shown that recognition is possible at all orientations with the baseline algorithm though there is a small drop in performance near 90 degrees (in front of the subject). Changes in orientation of 20 degrees between comparison sets significantly degrade recognition ability with this algorithm, although the effect is marginally better at orientations near side on. This property can be explained by the motion of the legs being scaled approximately by the cosine of the angle from the side on view.

Assuming that generating or recording a gallery at orientations 10 degrees apart is impractical, this imposes a severe constraint upon the utility of this algorithm. There may be ways of manipulating the silhouettes to reduce the number of galleries needed, it is possible there are silhouette features that are independent of orientation or that can be made more independent of orientation. In the next chapter this experiment is repeated with another algorithm and some simple ways to increase orientation independence are examined.

## Chapter 4

# Average Silhouette at Multiple Orientations

The previous chapter (Chapter 3) showed that the performance of the baseline algorithm degrades rapidly as the angle between the probe and gallery increases. Several questions are raised about these results, do the results based upon silhouettes generated with the SGAW truly reflect the performance of the algorithm in a real world scenario, and, is the baseline algorithm itself more prone to orientation effects than another?

The baseline algorithm is quite ‘raw’, essentially comparing silhouettes directly, whereas an algorithm that calculates and compares more abstract properties of the silhouette sequence may have more tolerance for orientation variations. Additionally, there may exist a method of alleviating the error introduced by orientation or of working around it. Complete independence from orientation may be impossible or impractical but it may be possible and practical to reduce the error associated with it to a tolerable and viable level.

There are many algorithms we could experiment with (see Section 2.3), most use feature vectors considerably smaller than the raw representation of the baseline, and are silhouette based. Choosing a silhouette based algorithm will allow easier comparison with the baseline results and implies a wider application of any conclusions. Within the silhouette based algorithm category there are a wide range of approaches, some similar to the baseline and others very different. A very different approach is appealing, it would be interesting to see the effect of orientation upon a radically different algorithm such as the symmetry of the silhouette (Hayfron-Acquah et al., 2001), however a similar algorithm allows easier analysis of the source of any variation in results. In this chapter we build upon the results of the previous chapter with an algorithm similar to the baseline algorithm, in the next chapter a different type of algorithm is examined.

A locally developed silhouette based algorithm is examined over a range of orientations. The Average Silhouette algorithm (Veres et al., 2004) is powerful yet easily interpretable, similar to the baseline algorithm but with a smaller feature. Essentially, the algorithm calculates the average of a sequence of silhouettes of a subject walking. Where the silhouette doesn't differ greatly, for example at the head and torso, the average value is large. Where parts of the body are moving the average silhouette has a value smaller. The average pixel intensity values are used as co-ordinates in a large dimensional space, this feature vector has as many dimensions as there are pixels in a single silhouette. Many pixel values will be correlated, PCA is then used on the feature vector gallery to project the large feature vectors into a lower dimensional space.

A simpler algorithm than that described in Veres et al. (2004) was implemented; PCA based dimensionality reduction was not used, the 'large' feature vectors were not computationally impractical. The City Block distance between feature vectors was implemented rather than the Mahanobis distance. There are many feature processing methods and comparison functions, and a multitude of combinations that could be applied. Investigating each combination regarding orientation performance is impractical and redundant as they are all ultimately reliant on the gait algorithm. Therefore we use a simple feature processing (normalisation) and K-NN for investigation of different algorithms. It is expected that more advanced feature processing and comparison will improve the performance algorithms.

## 4.1 The Average Silhouette Algorithm

When applying this algorithm it is necessary to extract a contiguous set of silhouettes one stride in length, alignment of the beginning and end of the sequence within the stride cycle is not necessary. Once the period of the subject's gait is calculated any period length sequence can be used. The silhouettes must be of the same scale and aligned in the images before the algorithm is applied.

Each frame in sequence  $S$  is  $x$  pixels wide and  $y$  high,  $S$  is  $T$  long where  $T$  is the rounded period of the subject's gait. An individual pixel of the silhouette sequence is denoted as  $S(i, j, t)$  where  $i \in 1 \dots x, j \in 1 \dots y$  and  $t \in 1 \dots T$ . First the centre of mass  $i_{cm}, j_{cm}$  of the original silhouette  $S_O(i, j, t)$  is found (assuming a foreground pixel is 1 and background is 0);

$$i_{cm}(t) = \frac{\sum_{i=0} \sum_{j=0} i * S_O(i, j, t)}{\sum_{i=0} \sum_{j=0} S_O(i, j, t)} \quad (4.1)$$

$$j_{cm}(t) = \frac{\sum_{i=0} \sum_{j=0} j * S_O(i, j, t)}{\sum_{i=0} \sum_{j=0} S_O(i, j, t)} \quad (4.2)$$

Equation 4.3 shows how the original silhouette of the sequence  $S_O$  are rescaled to  $p$  pixels high and the individual image centre of mass aligned at the image centre  $i_c(t), j_c(t)$ .

$$S(i, j, t) = S_O \left( i_c + \frac{p}{j_{max} - j_{min}} * (i - i_{cm}(t)), j_c + \frac{p}{j_{max} - j_{min}} * (j - j_{cm}(t)) \right) \quad (4.3)$$

A feature vector  $V$  is generated that describes the subject in a video sequence  $S$ ,  $V$  is also  $x$  pixels wide and  $y$  high, and is generated with:

$$V_{i,j} = \frac{1}{T} \sum_{t=1}^T S(i, j, t) \quad (4.4)$$

When implemented  $V$  can be a flat  $x * y$  long vector, the  $x$  by  $y$  arrays are presented here for ease of visualization. In order to compare two feature vectors the distance  $d$  between feature vectors  $V^1$  and  $V^2$  is calculated using :

$$d = D(V^1, V^2) = \sum_{i=1}^x \sum_{j=1}^y |V_{i,j}^1 - V_{i,j}^2| \quad (4.5)$$

$d$  can be thought of as the sum of the difference of intensity values of the two feature vectors. This is the distance between two feature vectors or average silhouettes. It is apparent that:

$$D(V^1, V^2) = D(V^2, V^1) \quad (4.6)$$

Treating  $V$  as a point in an  $x * y$  dimensional feature space,  $D$  calculates the city block distance between two points not pixel positions. The Euclidean distance between two feature vectors;

$$d = D(V^1, V^2) = \sqrt{\sum_{i=1}^x \sum_{j=1}^y (V_{i,j}^1 - V_{i,j}^2)^2} \quad (4.7)$$

could be used, but the city block metric has more potential for experimentation. The sum of the difference in average silhouette values is more directly understandable than the Euclidean distance, plus, it is trivial to see that;

$$\sum_{i=1}^x \sum_{j=1}^y |V_{i,j}^1 - V_{i,j}^2| = \sum_{i=1}^x \sum_{j=1}^N |V_{i,j}^1 - V_{i,j}^2| + \sum_{i=1}^x \sum_{j=N}^y |V_{i,j}^1 - V_{i,j}^2| \quad (4.8)$$

Indeed, the total difference between two feature vectors is the sum of the difference of all its regions, however partitioned. Although not used here, this would allow the

assessment of the contributions of various regions of the average silhouette to the total difference.

## 4.2 Experimentation upon SGAW Data

This algorithm was applied to the SGAW database described in Section 3.1, the period of the subjects gait was calculated as before in Section 3.2 and a single period of silhouettes extracted. 4 instances of all 19 people at orientations of 0, 5, 10, 15...90 degrees to the normal view were processed and average silhouettes generated. Figure 4.1 has sample frames from person 5105, instance 1 with angle = 0, with the sequence average Figure 4.1(d). The data used was the same as that in the previous experiment (Chapter 3).

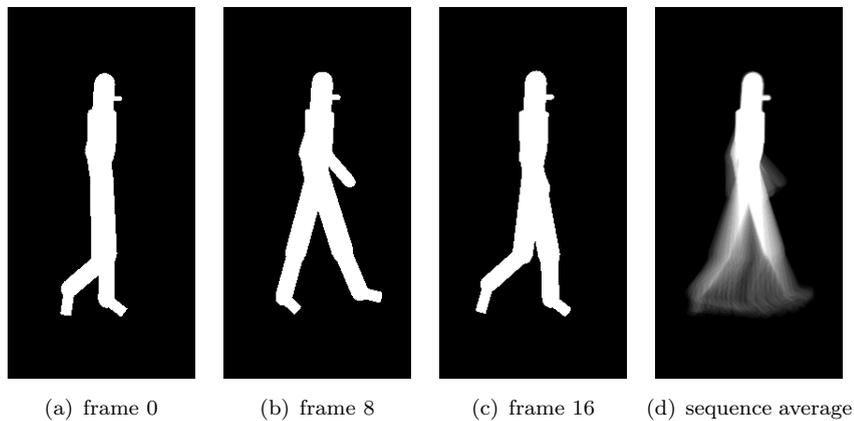


FIGURE 4.1: 3 Sample frames from person 5105, instance 1, angle = 0, with the sequence average

Average silhouettes were generated for each instance of a person (Figure 4.2(a) to (d)), ideally these should be identical. The 19 people had their average silhouettes generated (Figure 4.2(e) to (h)). The more distinct the average silhouette the easier recognition will be. Feature vectors are expected to change significantly with large changes in orientation, the shape of someone walking toward you is different to that of someone walking normal to your view, this is expressed in the feature vectors shown in Figure 4.2(i) to (l).

Before attempting recognition, the intra and inter subject variability of the average silhouettes was examined with respect to how it changes with orientation. For each subject, the mean silhouette at 0 degrees was created by averaging the 4 instances, then the distance between each of the four instances and the mean was found and plotted in Figure 4.3. This gives an indication of the internal variability of a subject. The results imply that the instances of some subjects (e.g. 5155) are more tightly clustered than others (e.g. 5170) and that the differences range between 50 and 250.

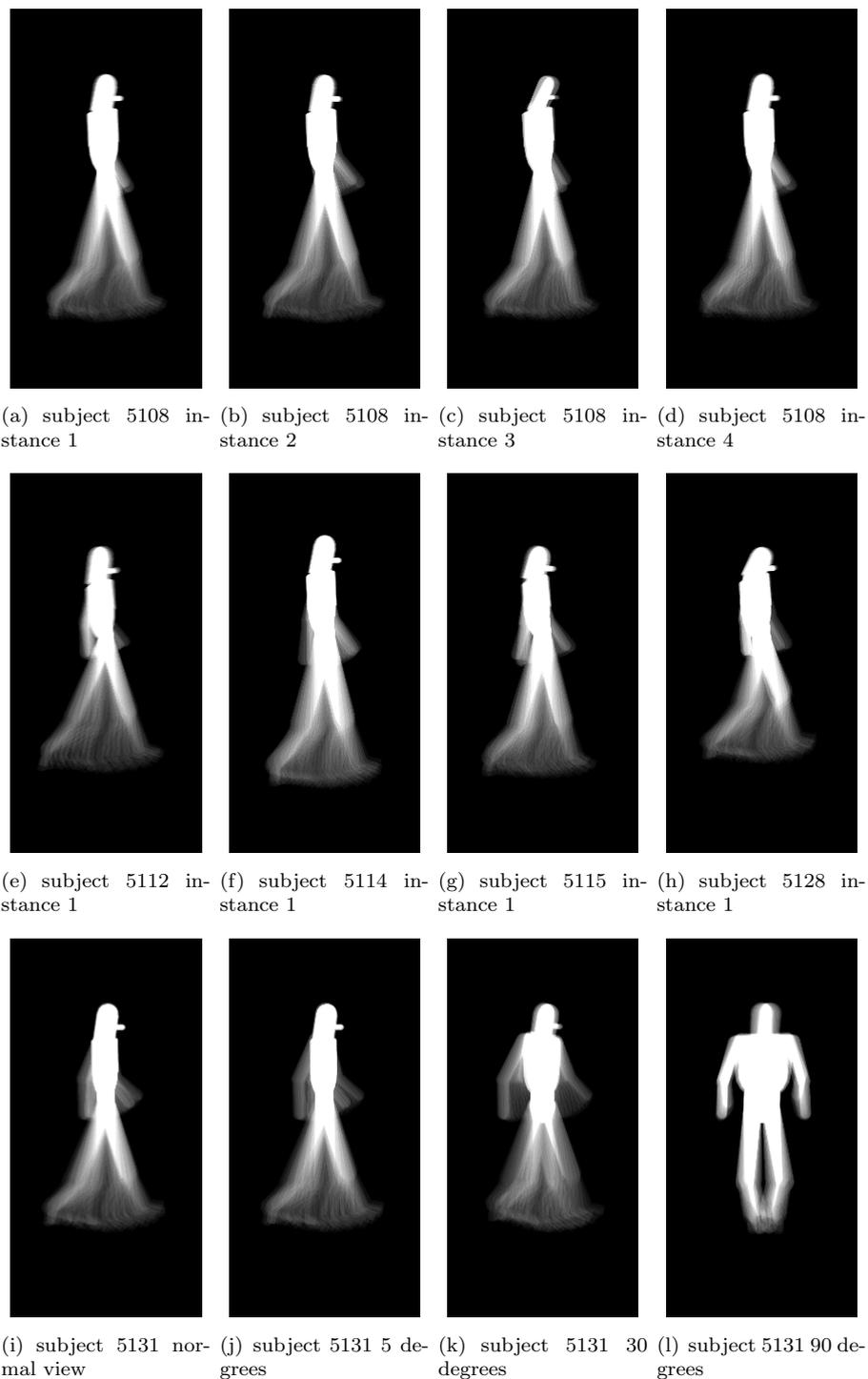


FIGURE 4.2: Example feature vectors. 4 instances of the same person (a-d), 4 different people (5112,5114,5115,5128) e-h and 4 different orientations (0,5,30,90) i-l.

Figure 4.3 begs the question, do these differences between feature vectors change with orientations? If this was the case, the inter or intra subject variance changed may be more amenable to recognition at a specific orientation.

Figure 4.4 shows the distance between the instances of subject 5105 and the mean of

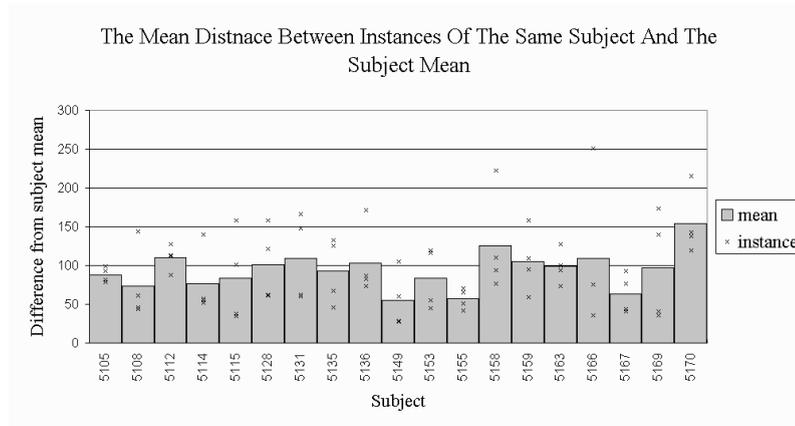


FIGURE 4.3: For orientation 0, the difference between the 4 instances and mean of each subject. The mean deviation of each person is similar but there is a significant variance.

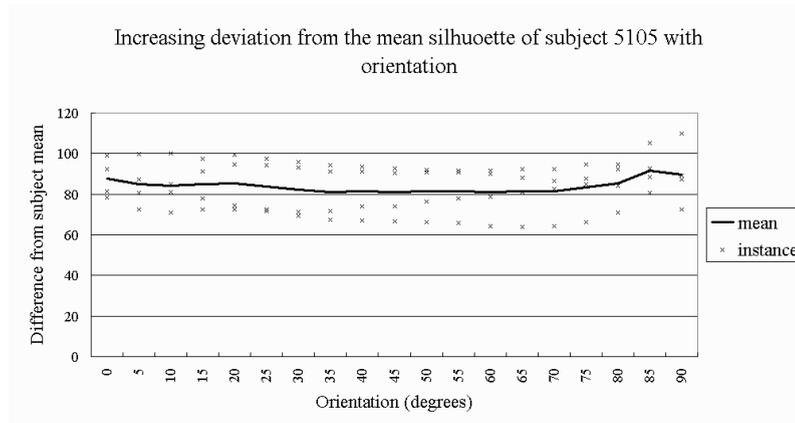


FIGURE 4.4: The differences of the 4 instances of subject 5105 from the mean of 5105 at 5 degree intervals to 90 degrees.

the four instances at each orientation, however it does not change significantly with orientation. Other subjects displayed similar constancy across silhouette orientation and therefore this was not explored further.

Figure 4.5 is the counterpart to Figure 4.3 and compares the mean of all instances of each subject (calculated for Figure 4.4) with the mean of all subjects. The variation between each subject mean and the global mean is between 200 and 600. This is more than that in Figure 4.3 and suggests recognition is possible.

Another way to view the differences between feature vectors is with a confusion matrix. Figure 4.6(a) contains the information in Figure 4.5 and Figure 4.3. The matrices for all orientations are in Section B.1. Because  $D(V^1, V^2) = D(V^2, V^1)$ , the matrices are symmetric over the diagonal. The relatively small differences recorded along the diagonal indicates that the intra subject distances are smaller than the inter subject differences and that recognition should be relatively successful. The values of all three matrices have been scaled by the same amount, and the off diagonal values in the 90 degree matrix

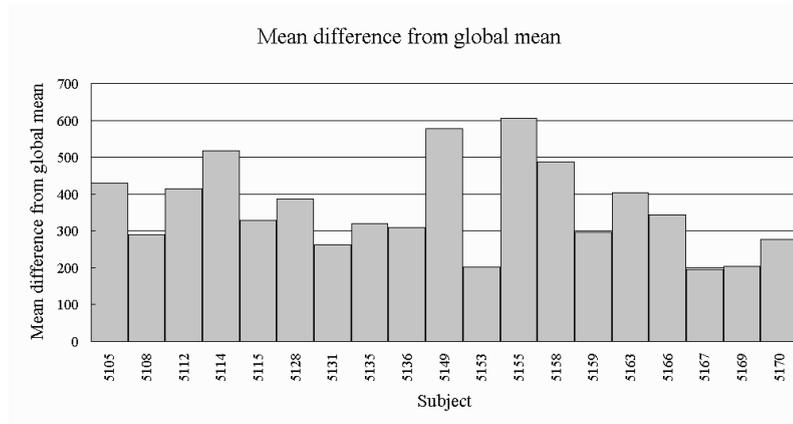


FIGURE 4.5: The difference of each subject from the global mean feature vector.

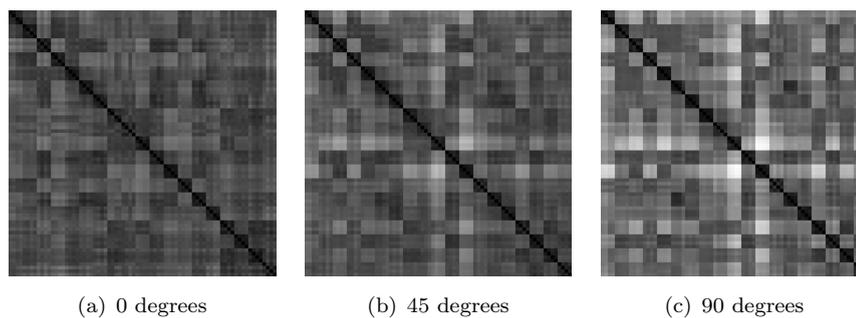
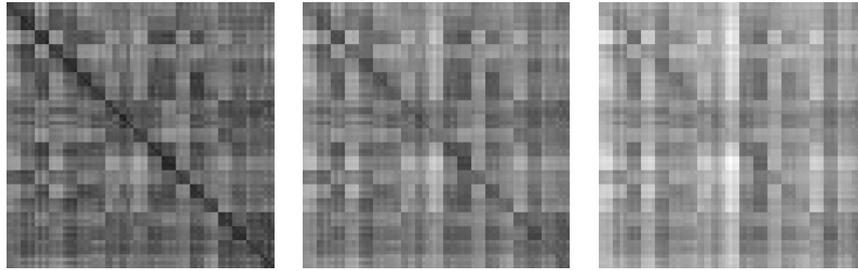


FIGURE 4.6: The confusion matrix of all runs of all subjects at 0,45 and 90 degrees. Each pixel row and column represents one instance of a subject. The 4 by 4 boxes of a similar shade show the internal consistency of each subject. The 4 by 4 dark boxes along the diagonal indicates there is little difference between instances of the same subject and a large difference between subjects.

are brighter (inter person differences are greater), this suggests that the feature vectors recorded at 90 degrees may be more unique than those captured at 0 or 45 degrees.

Figure 4.7 shows an effect of orientation upon the algorithm, similar to Figure 4.6 it compares every instance with every instance and plots the similarity of the feature vectors generated, however for this experiment, instances from different orientations are compared. Figure 4.7(a) shows the confusion matrix representing a comparison of all subjects at 0 degrees with those at 10, along the diagonal an instance is still compared to itself but it is an instance of that walk from a different view. The apparent confusion increases as the difference in orientation increases (Figure 4.7(c)), this is visible as the diagonal becomes lighter relative to the other sequences and the general lightening of the matrix. The confusion matrices for all orientations, not just 10, 20 and 30 degrees, are in Appendix B.

Differences in orientation may produce larger changes in the feature vectors at some orientations than others. Figure 4.8 shows the mean difference between adjacent angles (0 and 5, 5 and 10 and so on) for all instances of all people. There is less change in



(a) 0 degrees vertically, 10 horizontally (b) 0 degrees vertically, 20 horizontally (c) 0 degrees vertically, 30 horizontally

FIGURE 4.7: The confusion matrix of all runs of all subjects at 0 degrees vs other orientations.

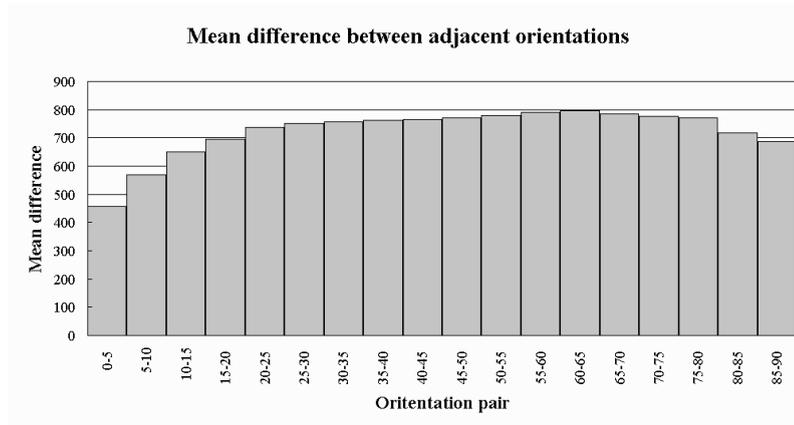


FIGURE 4.8: The mean difference between adjacent angles

feature vectors between low angles (views nearly normal) and high angles (nearly frontal views). The position of the extended feet are approximately in a plane normal to the camera view at 0 degrees, the horizontal position of the foot changes approximately with  $\cos \theta$  where  $\theta$  is the orientation. The change in the silhouette will be proportional to  $|\frac{d \cos \theta}{d \theta}|$  which is small at small  $\theta$ . The small change at large angles is due to the width of the body and the changing position of the shoulders. The shoulders are approximately in a plane parallel to the camera view axis, their position will change with  $\sin \theta$  and the change in silhouette will be proportional to  $|\frac{d \sin \theta}{d \theta}|$  which is small at large  $\theta$ . The difference between feature vectors will be greater when the difference in the subjects orientation is larger, Figure 4.9 confirms this.

Figure 4.10 shows the average similarity between all the angles for all instances of all people.

### 4.3 Recognition Experiments with the SGAW

For each orientation one of the four instances of each subject is used as the probe. The gallery (that the probe is compared with to find the closest match) consists of the other

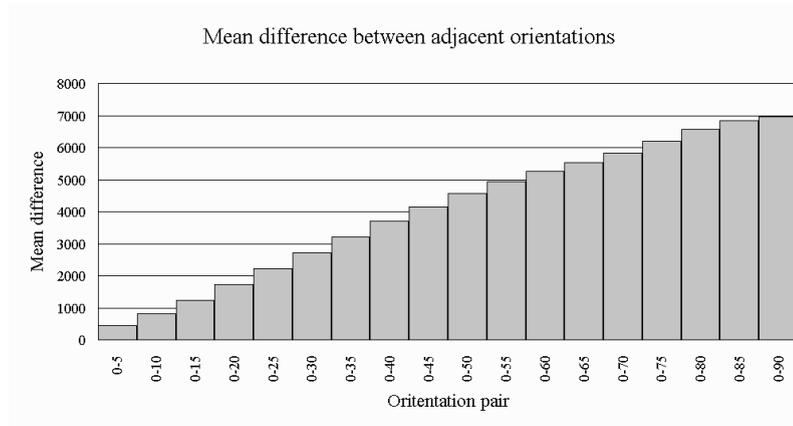


FIGURE 4.9: The mean difference between 0 degrees and the other angles

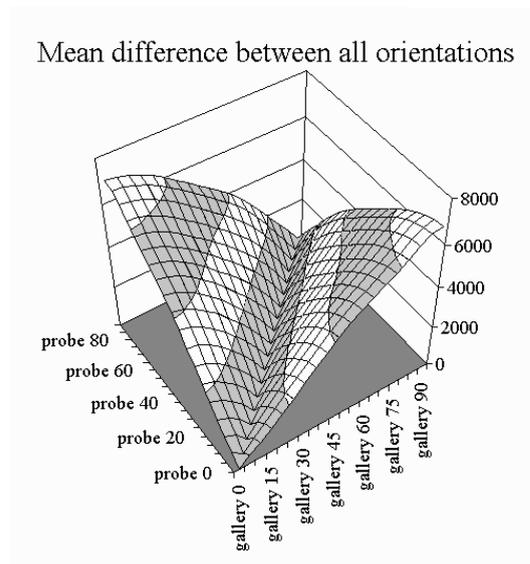


FIGURE 4.10: The mean difference between all angles

three instances of all subjects. For example if the probe instance is 2, the gallery consists of feature vectors from instances 1,3 and 4 from all 19 subjects, this ensures a probe is not compared against itself. The 4 recognition rates (probe instances 1, . . . ,4) are then averaged to generate the recognition rate for the experiment.  $K$  nearest neighbour with  $K = 3$  was used to decide identity.

### 4.3.1 Gallery at Probe Orientation

The first recognition experiment simulates the situation faced if either there is a gallery recorded at every orientation, or synthesizing an accurate gallery at an arbitrary orientation is possible. The feature vectors of the probe are extracted from silhouettes at the same orientation as the gallery.

Probe Instance	Orientation									
	0	5	10	15	20	25	30	35	40	45
1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
3	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Mean	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Probe Instance	Orientation									
	50	55	60	65	70	75	80	85	90	
1	100.0	100.0	100.0	94.7	94.7	94.7	94.7	94.7	94.7	
2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	
3	94.7	100.0	100.0	100.0	100.0	100.0	94.7	89.5	94.7	
4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	
Mean	98.6	100.0	100.0	98.6	98.6	98.6	97.3	96.0	97.3	

TABLE 4.1: Recognition rates for when the probe and gallery are at the same orientation.

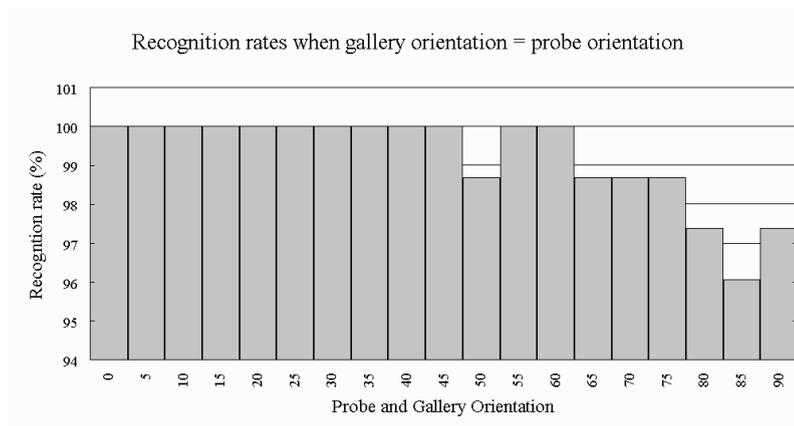


FIGURE 4.11: The recognition rate at each orientation, where, orientation of the probe = orientation of the gallery

Table 4.4 contains the recognition rates observed at each orientation with each set of instances as the probes. The table also contains the mean recognition rate at each orientation, this is plotted in Figure 4.11. With probes and galleries at the same orientation excellent recognition rates at all orientations are achieved, there is a small drop in performance at orientations close to 90 degrees (frontal views), although these recognition rates are still above 95%. There is an anomalous incorrect classification at 50 degrees, inspection of the underlying data reveals this a chance misclassification of a pathologically inconsistent subject (large intra subject variance).

If there is not a feature vector in the gallery at the orientation of the probe and synthesizing a gallery at the probes orientation is not possible, the probe and feature vectors at different orientations must be compared. This experiment compares probes from all orientations (0,5,...,90 degrees) with galleries at all orientations. For each combination

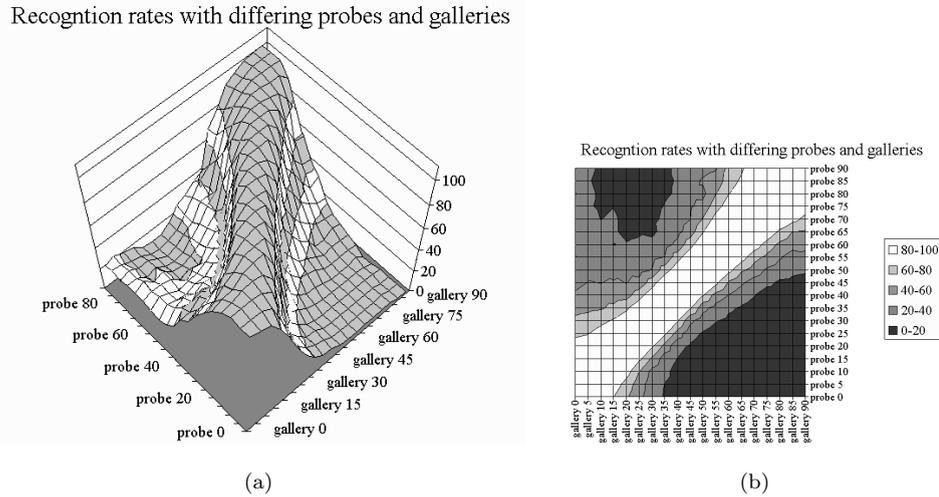


FIGURE 4.12: The recognition rate with galleries and probes at different orientations.

of probe and gallery orientation each of the 4 instances were used as probes and the corresponding instance removed from the gallery as in the previous experiment,  $K$  nearest neighbour with  $K = 3$  was used for deciding which gallery subject corresponded to the probe subject.

Figure 4.12 and Table 4.2 describe the recognition rates with probes and galleries at various orientations. The diagonal ridge of Figure 4.12 where the probe and galleries are at identical orientations corresponds to Figure 4.11. The ridge is wider where the orientation is close to 0 or 90 degrees. This may be because differences in orientation do not cause as large differences in the feature vectors at those orientations, as shown in Figure 4.8.

Recognition rates fall as the angle between probe and gallery increases, perhaps a subset of galleries can perform acceptable recognition at all orientations. Figure 4.13 shows the recognition rates when probes are compared to galleries at either 0,45 or 90 degrees depending on which they are closest to. This relies upon knowing the probe angle accurately, having the ability to record the three galleries and yet still exhibits a large drop in performance compared to Figure 4.11.

Finding the mean recognition rate for probes at all orientations  $(0, 5, \dots, 90)$  against a gallery  $g_\theta$  at orientation  $\theta$  gives a simple measure of the total recognition ability across all orientations. The gallery at 5 degrees had the highest mean recognition, Figure 4.14 shows the high recognition rate with probes within 15 degrees though it quickly drops off past 20 degrees. Encouragingly, the recognition rate is still 20.

Probe Orientation	Gallery Orientation									
	90	85	80	75	70	65	60	55	50	45
90	97.4	94.7	90.8	80.3	63.2	44.7	28.9	18.4	11.8	11.8
85	96.1	96.1	96.1	90.8	72.4	48.7	28.9	19.7	14.5	11.8
80	97.4	97.4	97.4	96.1	89.5	60.5	34.2	23.7	14.5	13.2
75	93.4	94.7	97.4	98.7	96.1	85.5	56.6	28.9	19.7	11.8
70	86.8	89.5	90.8	97.4	98.7	100.0	84.2	53.9	23.7	18.4
65	76.3	78.9	86.8	90.8	97.4	98.7	98.7	84.2	51.3	26.3
60	65.8	68.4	75.0	81.6	92.1	97.4	100.0	98.7	82.9	44.7
55	51.3	55.3	57.9	64.5	76.3	94.7	98.7	100.0	98.7	81.6
50	44.7	44.7	52.6	50.0	51.3	71.1	92.1	98.7	98.7	97.4
45	30.3	40.8	39.5	42.1	39.5	44.7	71.1	93.4	100.0	100.0
40	22.4	32.9	35.5	31.6	35.5	32.9	40.8	71.1	93.4	100.0
35	19.7	27.6	25.0	23.7	26.3	26.3	32.9	40.8	72.4	93.4
30	21.1	27.6	23.7	19.7	21.1	22.4	26.3	32.9	44.7	73.7
25	22.4	25.0	25.0	22.4	15.8	18.4	15.8	26.3	34.2	48.7
20	22.4	27.6	19.7	22.4	15.8	14.5	17.1	22.4	31.6	39.5
15	22.4	22.4	17.1	19.7	13.2	11.8	13.2	21.1	27.6	38.2
10	19.7	22.4	18.4	18.4	13.2	9.2	11.8	17.1	22.4	35.5
5	22.4	21.1	17.1	15.8	10.5	9.2	10.5	14.5	22.4	32.9
0	19.7	22.4	18.4	17.1	10.5	9.2	9.2	14.5	25.0	27.6
Probe Orientation	Gallery Orientation									
	40	35	30	25	20	15	10	5	0	
90	6.6	5.3	5.3	5.3	5.3	5.3	5.3	2.6	2.6	
85	6.6	6.6	5.3	5.3	5.3	5.3	5.3	2.6	2.6	
80	7.9	6.6	6.6	5.3	5.3	5.3	5.3	2.6	2.6	
75	7.9	6.6	6.6	5.3	5.3	5.3	5.3	3.9	2.6	
70	11.8	9.2	6.6	6.6	5.3	5.3	5.3	3.9	0.0	
65	18.4	11.8	11.8	7.9	5.3	5.3	5.3	6.6	0.0	
60	30.3	21.1	17.1	14.5	7.9	6.6	6.6	6.6	5.3	
55	46.1	32.9	22.4	18.4	13.2	7.9	9.2	9.2	7.9	
50	81.6	52.6	39.5	27.6	21.1	9.2	9.2	14.5	11.8	
45	97.4	85.5	57.9	43.4	31.6	25.0	17.1	15.8	14.5	
40	100.0	96.1	86.8	63.2	51.3	36.8	30.3	23.7	22.4	
35	100.0	100.0	100.0	89.5	67.1	55.3	42.1	31.6	35.5	
30	94.7	100.0	100.0	100.0	92.1	72.4	55.3	46.1	44.7	
25	78.9	93.4	100.0	100.0	100.0	94.7	76.3	59.2	57.9	
20	61.8	88.2	96.1	100.0	100.0	98.7	96.1	86.8	73.7	
15	44.7	72.4	88.2	96.1	100.0	100.0	98.7	98.7	92.1	
10	35.5	59.2	78.9	90.8	96.1	100.0	100.0	100.0	100.0	
5	42.1	48.7	68.4	82.9	93.4	96.1	100.0	100.0	100.0	
0	34.2	38.2	67.1	77.6	90.8	94.7	98.7	100.0	100.0	

TABLE 4.2: Recognition rates when the probe and gallery are at different orientations.

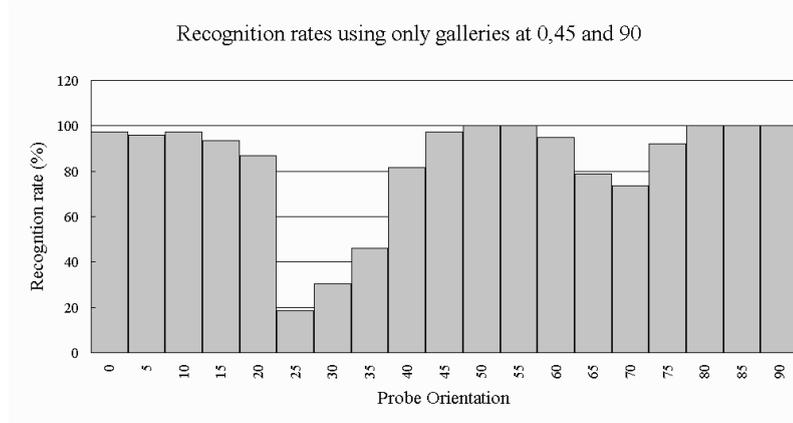


FIGURE 4.13: The recognition rate attainable if only galleries at 0,45 and 90 are available and probes are compared to the closest gallery. There are large drops in recognition when the the probe is far from one of the three galleries.

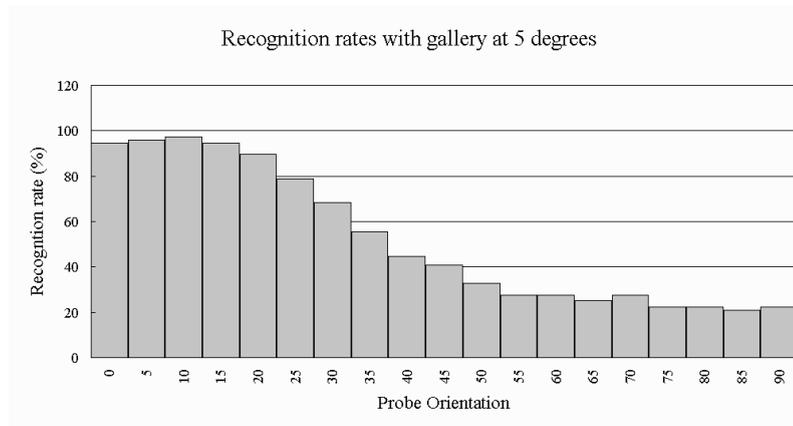


FIGURE 4.14: The recognition rate across all orientaions with a gallery at 5 degrees.

## 4.4 Recognition Rates with the SGAW data using Planar People

The planar person approximation (Kale et al., 2003a) may have some ability to correct for orientation. The average silhouette algorithm was modified and re-applied based upon the assumption that people are two dimensional. The feet, when maximally separated describe the furthest points from the centre of the body in the plane normal to the vertical axis. Following Kale et al. (2003a) people are modelled as being 2 dimensional in the plane described by the direction of motion and the vertical axis.

Equation 4.4 constructs feature vector  $V$  from sequence  $S$ :

$$V_{x,y} = \frac{1}{T} \sum_{t=1}^T S_{x,y,t} \quad (4.9)$$

and Equation 4.10 which transforms image  $I$  with dimensions  $(X, Y)$  from orientation  $O_a$  to  $O_b$ :

$$I_{x,y}^b = I_{x',y}^a \quad (4.10)$$

where

$$x' = (X/2) + (x - X/2) * \frac{\cos(O_b)}{\cos(O_a)} \quad (4.11)$$

To transform feature vector  $V^a$  at orientation  $O_a$ :

$$V_{x,y}^a = \frac{1}{T} \sum_{t=1}^T S_{x,y,t} \quad (4.12)$$

to  $V^b$  at orientation  $O_b$ . The original silhouettes can be transformed using Equation 4.11:

$$V_{x,y}^b = \frac{1}{T} \sum_{t=1}^T S_{x',y,t} \quad (4.13)$$

but because

$$V_{x',y}^a = \frac{1}{T} \sum_{t=1}^T S_{x',y,t} \quad (4.14)$$

the feature vector itself can be transformed

$$V_{x,y}^b = V_{x',y}^a \quad (4.15)$$

The previous recognition experiment was repeated, comparing probes from all orientations (0,5,...,90 degrees) with galleries at all orientations.  $K$  nearest neighbour with  $K = 3$  was once again used for deciding which gallery subject corresponded to the probe subject. However, the gallery feature vectors were transformed to the orientation of the probe using Equation 4.11 and Equation 4.15.

Figure 4.15(a) and Table 4.3 describe the recognition rates with probes and galleries at various orientations once corrected. The diagonal ridge of Figure 4.15(a) is where the probes and galleries are at identical orientations, this corresponds to Figure 4.11 as no correction took place.

People are insufficiently planar for this approximation to allow for silhouette translation between orientations. Translating silhouettes imparts more errors than it reduces, as the orientation of a silhouette from the normal view increases, the error increases also. However, Figure 4.15 shows that for comparisons between near normal views, a small improvement is gained. Comparing silhouettes where both are at non normal views

Probe Orientation	Gallery Orientation									
	0	5	10	15	20	25	30	35	40	45
0	97.4	93.4	88.2	67.1	36.8	21.1	14.5	6.6	5.3	5.3
5	96.1	96.1	94.7	78.9	43.4	22.4	14.5	6.6	5.3	5.3
10	97.4	97.4	97.4	94.7	65.8	30.3	15.8	6.6	5.3	5.3
15	93.4	96.1	97.4	98.7	88.2	48.7	19.7	9.2	5.3	5.3
20	85.5	90.8	93.4	97.4	98.7	89.5	30.3	14.5	5.3	5.3
25	63.2	69.7	82.9	93.4	97.4	98.7	82.9	23.7	7.9	5.3
30	47.4	50.0	59.2	64.5	88.2	97.4	100.0	77.6	17.1	5.3
35	35.5	38.2	36.8	34.2	51.3	65.8	97.4	100.0	67.1	6.6
40	23.7	26.3	22.4	15.8	23.7	34.2	55.3	96.1	98.7	51.3
45	22.4	21.1	18.4	13.2	11.8	13.2	17.1	39.5	92.1	100.0
50	15.8	14.5	14.5	9.2	9.2	7.9	7.9	9.2	25.0	84.2
55	14.5	11.8	7.9	9.2	9.2	7.9	7.9	5.3	7.9	11.8
60	11.8	11.8	7.9	10.5	9.2	9.2	7.9	6.6	5.3	6.6
65	13.2	11.8	7.9	10.5	9.2	9.2	7.9	7.9	6.6	5.3
70	14.5	15.8	13.2	10.5	9.2	9.2	7.9	7.9	7.9	6.6
75	14.5	11.8	13.2	13.2	10.5	7.9	5.3	5.3	5.3	5.3
80	11.8	10.5	11.8	10.5	10.5	7.9	7.9	5.3	5.3	5.3
85	14.5	14.5	11.8	11.8	18.4	10.5	6.6	5.3	5.3	5.3
90	5.3	5.3	5.3	5.3	5.3	5.3	5.3	5.3	5.3	5.3

Probe Orientation	Gallery Orientation									
	50	55	60	65	70	75	80	85	90	
0	5.3	5.3	5.3	5.3	5.3	5.3	7.9	6.6	5.3	
5	5.3	5.3	5.3	5.3	5.3	5.3	7.9	6.6	5.3	
10	5.3	5.3	5.3	5.3	5.3	5.3	7.9	6.6	5.3	
15	5.3	5.3	5.3	5.3	5.3	5.3	7.9	6.6	5.3	
20	5.3	5.3	5.3	5.3	5.3	5.3	5.3	6.6	5.3	
25	5.3	5.3	5.3	5.3	5.3	5.3	5.3	6.6	5.3	
30	5.3	5.3	5.3	5.3	5.3	5.3	5.3	6.6	5.3	
35	5.3	5.3	5.3	5.3	5.3	5.3	5.3	6.6	5.3	
40	5.3	5.3	5.3	5.3	5.3	5.3	5.3	6.6	5.3	
45	39.5	5.3	5.3	5.3	5.3	5.3	5.3	6.6	5.3	
50	100.0	25.0	5.3	5.3	5.3	5.3	5.3	7.9	5.3	
55	71.1	100.0	14.5	5.3	5.3	5.3	5.3	7.9	5.3	
60	10.5	39.5	100.0	5.3	5.3	5.3	5.3	6.6	5.3	
65	5.3	9.2	18.4	100.0	5.3	5.3	5.3	6.6	5.3	
70	6.6	5.3	5.3	9.2	100.0	5.3	5.3	5.3	5.3	
75	5.3	5.3	6.6	5.3	5.3	100.0	5.3	5.3	5.3	
80	5.3	5.3	5.3	5.3	5.3	5.3	100.0	5.3	5.3	
85	5.3	5.3	5.3	5.3	5.3	5.3	5.3	100.0	5.3	
90	5.3	5.3	5.3	5.3	5.3	5.3	5.3	5.3	100.0	

TABLE 4.3: Recognition rates for when the probe and gallery are at different orientations but corrected with the planar people approximation.

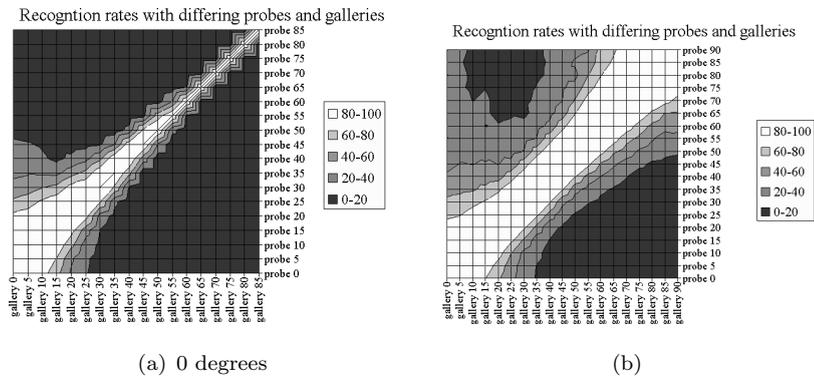


FIGURE 4.15: The recognition rate with galleries and probes at different orientations; (a) with a planar person approximation, (b) without any correction.

appears to add a combination of the error of both silhouettes to the similarity measure. It is to be expected the planar person approximation will not work for orientations far from normal, although it appears it is of marginal benefit for views close to normal.

	gallery	
	0°	22°
probe	0°	24%
	22°	74%

TABLE 4.4: Recognition rates between orientations.

## 4.5 Recognition with real subjects

The average silhouette algorithm was applied to the Southampton gait database (Shutler et al., 2002). The database has 2 views of over 100 subjects walking a linear track. One camera is situated normal to the track, the other approximately 25 degrees to normal. The database consists of sequences of silhouettes, background subtraction having been performed with the aid of chroma key extraction. 8 sequences of 50 people were used, two instances, walking left and right at two orientations, 0 (normal) and 25 degrees from normal. The silhouettes were cropped around their centroids to align them. To generate average silhouettes the algorithm described in Section 4.1 was applied.

Figure 4.16 shows four example average silhouettes. The head and body has a maximal value fading quickly to black at edges, the legs however are a blur due to their motion.

Based upon the experimentation on the SGAW data, there should be a drop in recognition rates between the two orientations. The SGAW recognition dropped by approximately 20.

A leave one out cross validation recognition experiment was performed, using one of the 4 sequences as the probe and the other 3 of all subjects as the gallery.

Table 4.4 contains the results of the recognition experiments; within an orientation the recognition rate is good but there is a significant drop in performance when using probes and galleries at different orientations. The drop more severe than Figure 4.12 would suggest. This may be due to several factors, the effect of perspective upon the subject, or that the SGAW camera used an orthogonal projection. Additionally there may be orientation specific noise in the silhouettes.

## 4.6 Conclusions

The results of the SGAW based experiments are similar to those of the previous chapter, this is of little surprise as the algorithms use silhouettes in a similar manner. Figure 4.12 suggests the algorithm can not accurately measure the similarity of subjects when the silhouettes are recorded at orientations that differ by more than 20 degrees.

Near the side on view the recognition rates between orientations is almost identical to that of the baseline algorithm. As the orientation of the views compared moves to 45

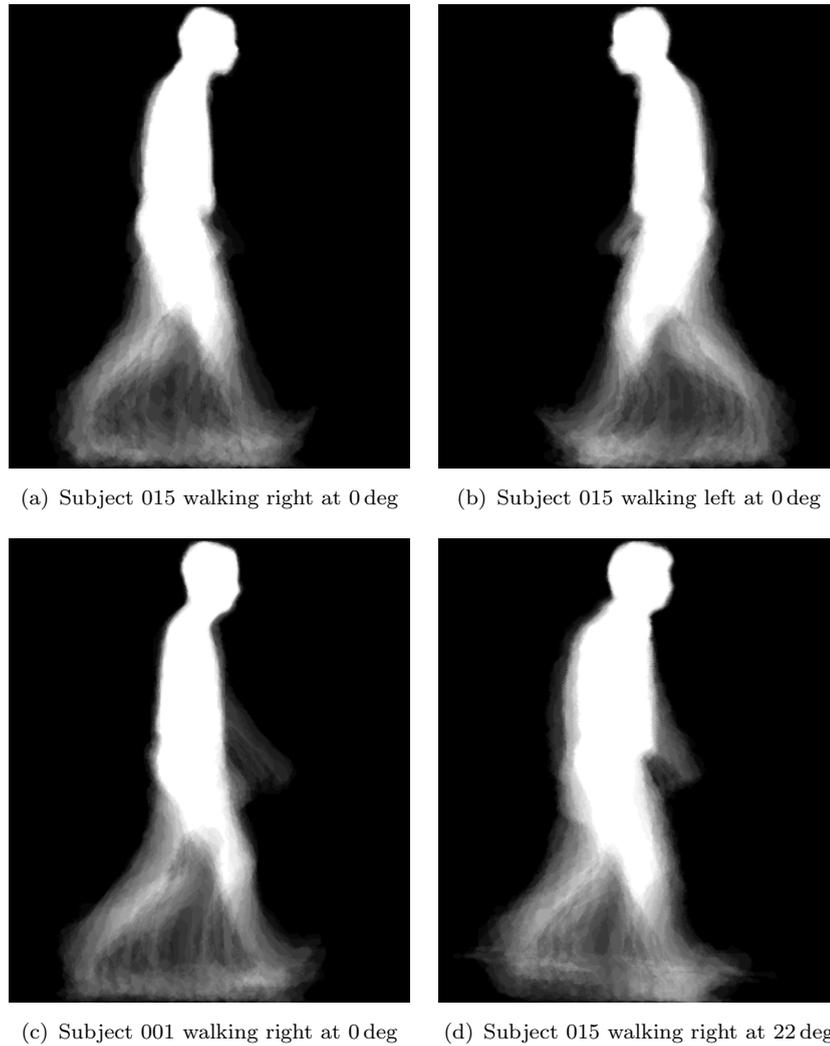


FIGURE 4.16: Example average silhouettes.

degrees and to the frontal view the performance is slightly better. With the baseline algorithm a drop of 20% using a 45 degree gallery occurs at 55 degrees, with the average image algorithm it is extended to 57 degrees. This extended operating range continues up to the frontal views. There are two factors that can explain this, the average image algorithm is less sensitive to the estimation of the phase and period of the walking subject and these properties are harder to accurately estimate at orientations further from normal, this would hinder recognition by the baseline algorithm at these orientations. The fact that the range over which recognition can be performed is greater closer to the frontal view than at 45 degrees may be because close to frontal the appearance of the upper body changes least with respect to orientation, and Veres et al. (2004) found the upper body and head had a large contribution to the recognition rate. Section 3.1 describes how the SGAW head was constructed and the precise and peculiar way SGAW head was drawn.

Similarly to the baseline algorithm, probes are ease to recognise with galleries at an

orientation closer to side on. This may be due appearance of the legs changing approximately with the cosine of the angle from side-on.

The planar person approximation gave no benefit to recognition with this algorithm. Comparing Figure 4.15 and Figure 4.12 it is apparent the error it introduces is at least as large as that it compensates for.

A similar effect to that observed in Figure 4.12 can be seen with real life data in Table 4.4. This confirms the effect described with SGAW is real and suggests the real silhouettes affected by orientation are more challenging than those simulated with SGAW. It is expected that the performance of the baseline algorithm to be worse than that of the average silhouette. As the a person passes the camera they are constantly changing in orientation with respect to the camera. The average image is then an amalgamation of images over a range of views. The baseline feature vector represents ordered frames beginning at a specific orientation as all feature vectors must start at the same part of the gait cycle.

Veres et al. (2004) found the upper body and head had a large contribution to the recognition rate. Section 3.1 describes how the SGAW head was constructed and the precise and peculiar way SGAW head was drawn, this could be a large factor in the high recognition rate recorded.

Additionally, noise present in the real silhouettes from clothing, irregular body shape, camera and perspective effects would all have contributed to the increased difficulty in recognising real world subjects.

The average silhouette algorithm and the baseline algorithm are similar, and use the silhouettes as a whole. The effect of orientation upon a different algorithm, one that uses a multitude of properties of the silhouette's sequence, may behave differently, and some properties of a silhouette sequence may be more independent of orientation than others.

## Chapter 5

# Orientation Independent Properties of Moments.

The preceding two chapters have shown the performance of silhouette based algorithms deteriorating rapidly as the difference in orientation between compared sequences increases. Are there measurements of silhouettes that are useful for recognition and independent of orientation? Height of the silhouette is independent of orientation, and the head is approximately rotationally symmetric around the rotational axis, these are however relatively weak biometrics.

In this chapter, the effect of orientation upon a moment based biometric is explored. Moments have already been successfully applied to gait recognition (Shutler and Nixon, 2001; Little and Boyd, 1998). Using moments of silhouettes, many properties of silhouettes can be calculated. Moments have a link to Fourier descriptors and a reconstruction theorem. Analysis of the calculated properties and their individual performance with respect to changes in orientation may reveal some that are independent of orientation.

### 5.1 Moment Generation

The use of moments on optical flow (Little and Boyd, 1998) or Zernike moments extended to include velocity (Shutler and Nixon, 2001) have been used successfully for recognition and shown to perform better than Cartesian moments. For this set of experiments Cartesian moments were adequate as it is their properties with respect to orientation under investigation rather than their absolute discriminative ability.

An image can be uniquely described by a set of discrete Cartesian moments  $M$  where  $M = [m_{0,0}, m_{0,1}, m_{1,0}, \dots, m_{p,q}]$ , and  $m_{p,q}$  is defined as:

$$m_{pq} = \sum_{x=1}^M \sum_{y=1}^N x^p y^q P_{xy} \quad (5.1)$$

Where  $M$  and  $N$  are the image dimensions,  $p$  and  $q$  are non-negative integers, and  $P_{xy}$  is the pixel value at  $(x, y)$ .  $m_{p,q}$  is a  $(p + q)$ th order moment. Each moment  $m_{p,q}$  describes a property of the silhouette, and an infinite series of Cartesian moments is required for a complete description of an image. In practice, high order moments tend to represent the high frequency noise present in the image. A subset of moments, up to a certain order, is taken from each silhouette in the image sequence and concatenated into a feature vector.

Cartesian moments are not independent under affine transformations, translation is easily accounted for by translating the silhouettes to their centroid before moment generation. There are no scale changes when using data generated by the SGAW and sober walking people are naturally rotationally aligned vertically.

To generate a feature vector an input sequence of silhouettes was clipped to one stride in length and the edges found with the union of the absolute values of convolutions of the silhouettes with horizontal and vertical  $[1, -1]$  masks. Calculating moments is a computationally intensive task, however by only processing the edge co-ordinates it is less demanding than the silhouettes themselves.

Moments of up to the 15th order were calculated from the edge of the silhouette of each frame, higher order moments were found to be extremely noisy. Inspecting moments showed they change smoothly between frames so used linear interpolation to down sample each moment to 25 frames.

The SGAW described in Section 3.1 was once again used to generate silhouettes, however the sample size was reduced to 3 sequences of 10 people at orientations between 0 and 90 (side on and frontal views) degrees at 5 degree intervals, due to the demanding nature of moment generation. The centroid of the edge co-ordinates were translated to the origin, scale and rotation remaining unchanged. Moments were calculated for each frame and concatenated into a single 3400 long feature vector (25 frames \* the 136 moments up to the 15th order). Each moment was normalized through all frames to  $0 \leq m_{pq} \leq 1$ , thus representing each sequence as a point in a 3400 unit feature space. This normalisation was necessary for comparing moments of different orders, an order 10 moment is much larger than an order 2 moment. By normalising each moment through the gait cycle, how the moment changes rather than its absolute value becomes its distinguishing feature.

Figure 5.1 shows an example of moment (2,0) of frame 8 changing through 90 degrees. A high value does not necessarily mean a large value of the moment, rather, a large value compared to the other values in the sequence. Moment (2,0) is the square of the x values of the edge pixels, frame 8 is near the part of the sequence where the legs are

furthest apart. As the angle is increased, the value of this moment relative to the rest of the sequence decreases due to the smaller projected motion of the legs. In this instance the near side leg is also the closest, therefore the legs appear to cross as the orientation increases. Suddenly, as the orientation approaches  $90^\circ$  the legs uncross, this causes the noise and spike in Figure 5.1.

## 5.2 Analysis of moment variance.

The Shapiro-Wilk test for normality (Shapiro and Wilk, 1965) was performed upon the moments generated and it was found that none of the moments were normally distributed. The null hypothesis of the test is that the samples are taken from a normal distribution, using  $P < 0.05$  we reject the null hypothesis. Parametric methods such as Analysis of Variance can not be used to analyse the moment data.

To explore how moments change with orientation and subject the Kruskal Wallis test was performed. The Kruskal Wallis test is a non-parametric ANOVA based upon the ordered rank of the observations. The test was performed twice for each moment, first with the moments grouped by subject and then by orientation. This is similar to 2D ANOVA but without the terms corresponding to the interaction of subject and orientation. The two null hypothesis being tested are:

1. The probability for the null hypothesis  $H_A$ , that samples from different orientations are drawn from the same population
2. The probability for the null hypothesis  $H_B$ , that samples from different subject are drawn from the same population

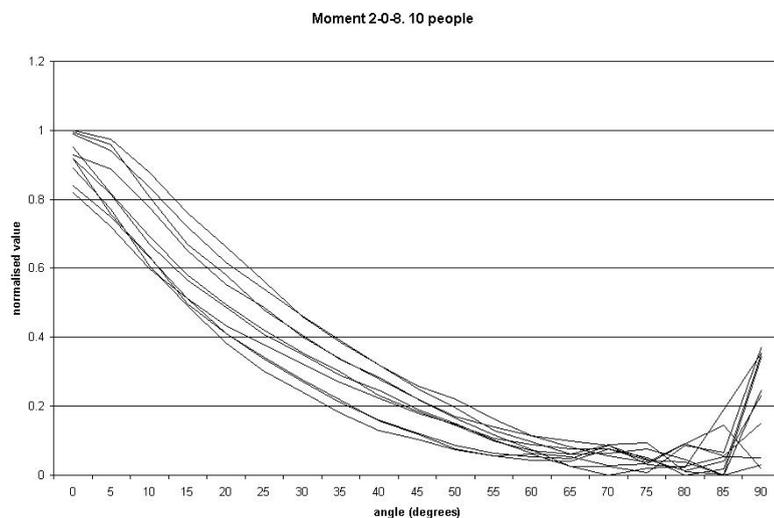


FIGURE 5.1: Moment (2,0) of frame 8 over ninety degrees plotted for 10 subjects. The moment appears to change with person identity but more with orientation.

Using these values we can explore the usefulness of each moment. If HA is false the moment is not independent of orientation. If HB is true the moment varies randomly with subject ID.

3263 moments were tested, with a significance level of 0.05. 28 were found to have significantly different medians, and of those 28, just 11 that did not also vary with orientation. Upon inspection it was found that these moments were reacting strongly to the width of the thighs vs the width of the chest when the legs are passing and approximately vertical. Unfortunately this is a feature of the SGAW and when present in real life, not a feature of gait.

### 5.3 Identification of Subject

A subject recognition experiment was performed using the same 3 instances of 10 people walking at intervals of 5 degrees between 0 and 90. Each instance was used as the probe in turn with the remaining two being used as the gallery, however assuming that the orientation was known, each classification was between 20 sets of moments (2 instances of 10 people at a known orientation). The all galleries plot of Figure 5.2 shows the recognition rate at each orientation.

The pre-requisite of recording a gallery near every possible orientation makes such a system impractical. An alternative to recording a large number of galleries is to compare moments calculated from silhouettes captured at different orientations. The recognition experiment was reformulated using galleries at intervals of 10 (0,10,20,...),15 (0,15,130,...),30 (0,30,60,90) and 45 degrees (0,45,90). Probes were classified using the gallery closest to their orientation. Figure 5.2 shows the recognition rates at all orientations using galleries at 0,45 and 90 degrees compared to using galleries at all 5 degree intervals. Probes from orientations furthest from the orientation of a gallery were unsurprisingly classified less accurately.

The error introduced by comparing moments calculated from silhouettes captured at different orientations may be alleviated if moments could be transformed between orientations. There are two possible places for such a transformation; either transforming the probe to the orientation of a gallery or transforming the gallery to the orientation of the probe. A practical system will most commonly have higher quality multiple cameras available during enrollment rather than capturing the probe.

To improve recognition of probes from orientations not represented in the gallery a virtual gallery was created by interpolating moments for each subject between available galleries. Figure 5.2 shows a small improvement in recognition using interpolated galleries. Figure 5.3 plots the average recognition rate across all orientations as a function of the interval between galleries. Using an interpolated gallery increased recognition rates

compared to using the nearest gallery but inaccuracies in the interpolated galleries still limited the size of the interval between recorded galleries. Using interpolated galleries allows the use of galleries recorded at 10 degree intervals with no loss in recognition.

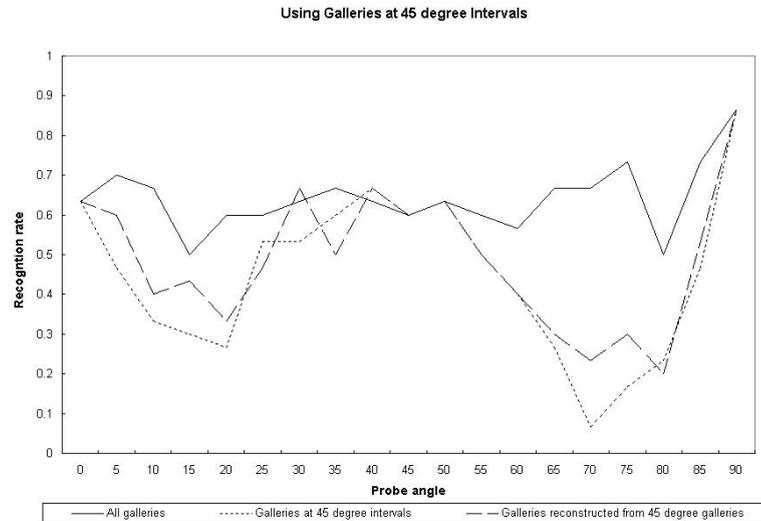


FIGURE 5.2: When only 3 galleries at 0,45 and 90 degrees are used, probes from intermediate orientations have a reduced recognition rate when compared to one of the 3 galleries. If a pair of galleries is interpolated to generate a virtual gallery at the same orientation as the probe, recognition at unrepresented orientations is improved.

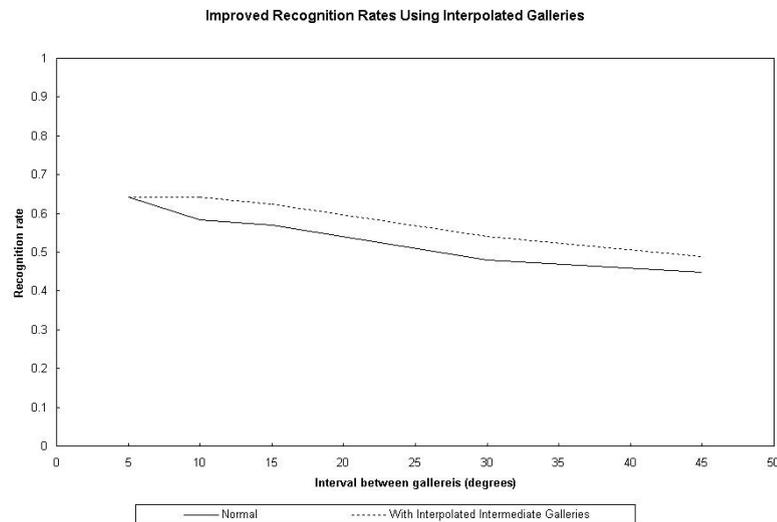


FIGURE 5.3: Average recognition rates are reduced when the spacing between the galleries is increased. Interpolating between galleries marginally improves recognition. Note that the probes are spaced at 5 degree intervals so the first point represents having a recorded gallery at the orientation of each probe.

Moment  $(2,0)$  of frame 8 only changes approximately linearly with orientation when the changes in orientation are small, see Figure 5.1. Consequently, linear interpolation between large changes in orientation is likely to fail with this moment. Inspection of other moments shows a similar pattern, to demonstrate that it is the subject dependant

change with orientation rather than the simple interpolation, the following experiment was performed.

Five subjects were used to generate an average set of moments for every angle (0–90 in 5 degree intervals). The average moments were subtracted from each gallery and probe. Figure 5.4 is a plot similar to Figure 5.2 but as only five people were used, recognition rates are higher. Subtracting the orientation average moment offered little improvement in recognition suggesting it is the interaction of person identity and orientation rather than the nonlinear change in moment with orientation that is causing the degraded recognition, i.e. the silhouettes of different people changing differently across different orientations. Looking at Figure 5.1, if a subject’s moment was consistent relative to the other subjects, average moment subtraction would normalise any moment.

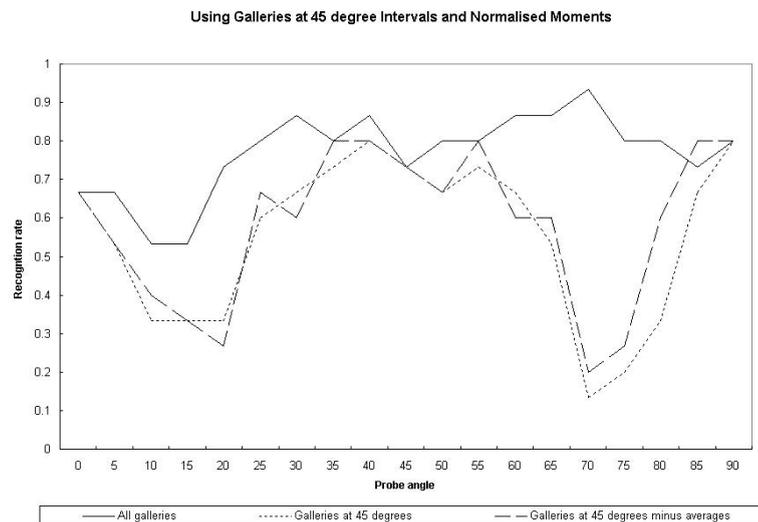


FIGURE 5.4: Five subjects were used to generate average moments for each angle and five were used for the experiment. Normalising the moments of each person using the averages gave a marginal improvement in recognition.

One more recognition experiment was performed several years after the original investigation into moments above. Due to advances in computing power it was possible to compute the recognition rates between all 4 runs of 19 people at the full range of orientations. This allows direct comparison with Figure 4.15 and Figure 3.9. An identical round robin leave one out recognition experiment was performed, the results can be seen in Figure 5.5.

Moments appear to have very little independence to orientation relative to the baseline algorithm or the average image algorithm, the performance dropping by 50% over difference in orientation of 10 degrees. Also it appears the greater tolerance to differing orientation is near the frontal views. At these views the upper body changes least and the sideways rocking of the body is most easily viewed.

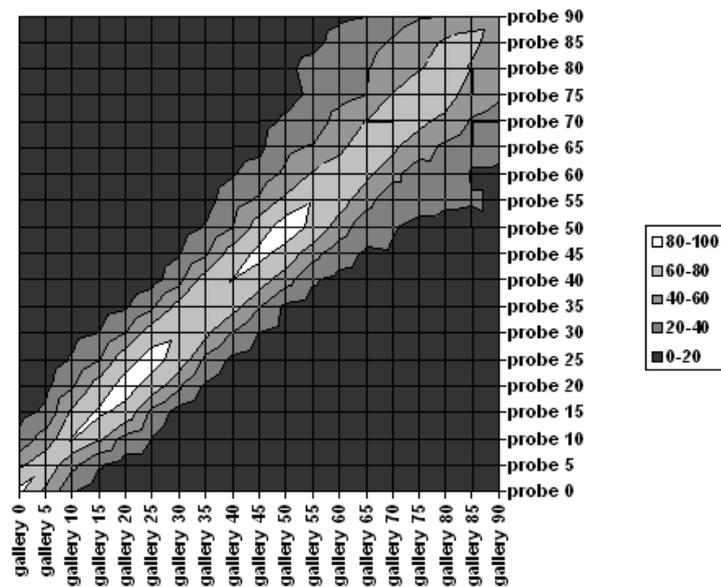


FIGURE 5.5: The recognition rate between probes and galleries at all orientations.

## 5.4 Determining the Orientation

Moments appear to have some ability to discriminate between orientation, a small experiment was conducted to examine whether orientation could be determined from moments of a silhouette. Using the same data as before, an almost identical recognition experiment was performed, all instances were in turn used as probes and compared to a gallery consisting of the other two instances of all subjects at all orientations. The estimated orientation being that of the gallery instance it is closest to.

To show how moments change with orientation Figure 5.6 was generated. For each pair of orientations the the distance between every run of every subject was found and normalised. Similar orientations have similar moments and the larger the change in orientation the larger the difference. Orientations near the front (90 degrees) appear to be the least similar to other orientations.

The subjects' identity was not used, therefore each classification was between all 10 people at 19 orientations. Figure 5.7 plots the match rate as a function of acceptable error, estimation of the orientation was successful and confirms the results in Section 5.2 that there is significant orientation information in the moment sequences despite irregularities introduced by subjects.

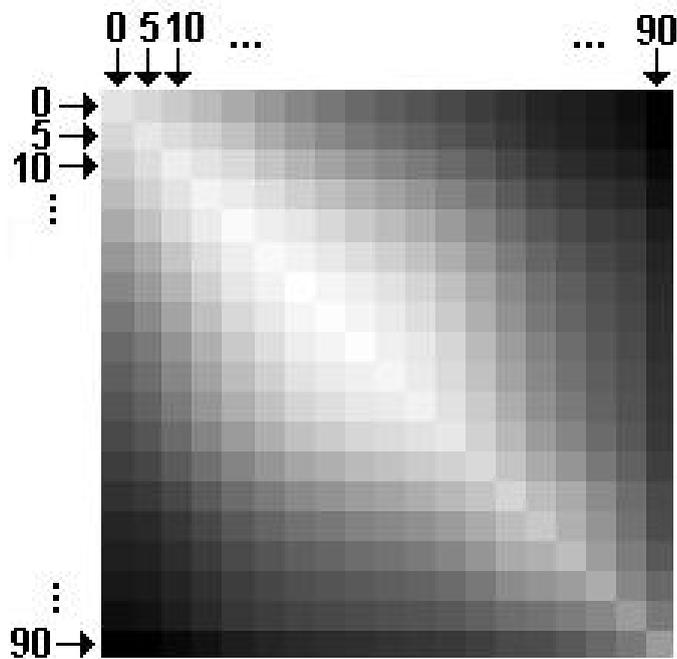


FIGURE 5.6: The similarity of moments from different orientations. The sum of the distance between every run of every subject between all combinations of orientation. Each square shows the sum of the distances of all runs rescaled to a greyscale. White is the minimum and black the maximum.

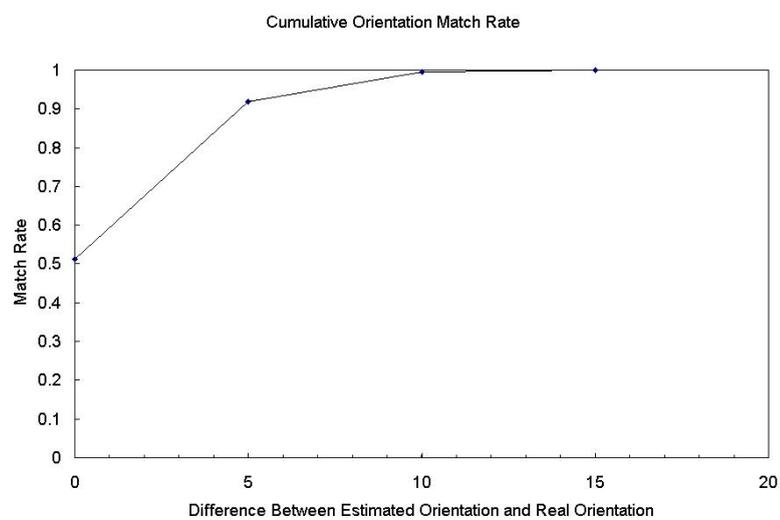


FIGURE 5.7: Only 50% of the orientations were recognised correctly, however 92% were estimated within 5 degrees of the real orientation.

## 5.5 Conclusions from Moment Experiments

The results of the baseline algorithm showed it breaking down when comparing silhouettes at orientations greater than 20 degrees ( Figure 3.9). The average image algorithm fared little better, and the limited use of approximating people as planar objects was demonstrated ( Figure 4.15 ). In this chapter interpolating between galleries and using a pre-recorded mean moments failed to yield significant improvements. The between orientation recognition was shown to be poor (Figure 5.5), and analysis of the moment medians failed to identify a subset of more useful moments.

The problem encountered in performing recognition using galleries at different orientations to the probe is due to the difficulty of generating a virtual gallery at unrepresented orientations. It is not clear how a moment can be translated between orientations or if it is possible. Cartesian moments have a reconstruction theorem, therefore transforming a set of moments back into silhouettes is possible, thus if silhouettes could be translated, moments could be and vice-versa. Moments and silhouettes change in a complex subject dependant manner with orientation. It appears highly unlikely a purely silhouette based algorithm can be improved to be independent of orientation.

The previous chapters have shown that several silhouette based algorithms fail to recognise people when viewed at different orientations. All three methods (moment based recognition, the baseline algorithm and the average silhouette) use the silhouettes themselves or statistics pertaining to the silhouettes for recognition. It was found that the statistics calculated can not be used at orientations that differ greatly from those at which they were recorded, e.g. features captured from a subject walking approximately normal to the camera's view during enrolment will not be useful for recognition at greater than 15 degrees from normal. If many views of the subject are available this is not a problem, however, if there is not an abundance of data, it becomes necessary to transform feature vectors between orientations.

If a feature can be transformed, the accuracy of the transform and the range of angles it is useful over is unknown. However, through examination the silhouettes and a prior knowledge of the structure and motion of a walking human, it may be possible to infer how a feature may be translated between orientations and under what circumstances it will be accurate. For example, the maximal width of a silhouette changes with the orientation of the subject with respect to the camera. If a subject's legs moved in the same plane, with the maximal width  $M$  when measured at the normal view, the maximal width  $m$  at orientation  $\theta$  can be described with;

$$m \approx M \cos \theta \tag{5.2}$$

This equation is an approximation, people are modelled as 3D even if their legs do swing in the same plane. Moving to a more realistic model where the legs move in distinct planes is possible but more information or assumptions are required about the width of the hips as maximal width alone does not provide this information. If people were two dimensional in a single plane, the entire silhouette could be transformed between orientations as every part of the subject is always visible. Apart from error introduced by quantisation there would exist a holistic transformation that could approximate any view. Unfortunately people are not 2D and treating them as if they are does not work (Section 4.4).

Of course people are not two dimensional, when viewed from different orientations different parts will contribute to the silhouette; some parts will become occluded appearing to move out of view while others are revealed. There are three properties of a solid human body that affect the shape of a projected silhouette at different orientations, the cross section, relative position of the centre of rotation and occlusion.

The projected cross section of some shapes does not change significantly between views, Figure 5.8 contains cross sections of two simple shapes and the projected edge position as they are rotated. Body parts with a circular cross section will not appear to change between orientations (Figure 5.8(b)). The head and neck are approximately spherical in cross section, as are the arms and legs.

If the cross section is not circular the projected position of the edges will change (Figure 5.8(c)). It can not be known what the cross section of an arbitrary part of a silhouette is. At most, constraints can be placed upon a cross section although calculating a silhouette far from a recorded view remains difficult. Given that the cross section is unknown and it can not be known how the edge of a silhouette will change between orientations, linearly interpolating edge positions by using the mean edge position between views is a reasonable option. Figure 5.8(d) is approximately linear for small changes in orientation although at  $90^\circ$  and  $270^\circ$  there are large changes in gradient and the potential for the mean edge position to have a large error. This is in effect what was attempted by using moments captured at separate orientations and interpolating them to calculate moments at orientations in between.

If the centre of rotation is not at the centre of the body part, the relative motion of the body part will move the edge of the silhouette in addition to the effects of the cross section above. The shape in Figure 5.9(a) is similar to Figure 5.8(a) but the centre of rotation is not at the centre of the object. The resulting change in the projected edge positions (Figure 5.9(b)) changes slowly and predictably, if this were the only effect (if people were 2D) it could be compensated for. This effect is in addition to any effect of the cross section described above.

Parts of the body may occlude each other differently at different orientations. Figure 5.10 shows how sharp changes in edge position occur when body parts occlude each other,

linear interpolation between these orientations will be inaccurate, unfortunately this situation is most common in the lower body as the legs occlude each other. If the parts of the body are small and far from the centre of rotation the potential error is exacerbated (Figure 5.10(d)).

The observed changes in silhouettes will be a combination of all three factors, the shape of the cross section, its displacement from the centre of rotation and occlusion. Complicating the problem is the fact that different parts of the silhouette are undergoing different changes due to different factors, using a model to segment the silhouette into body parts so that local factors can be accounted for is not possible in a holistic silhouette based algorithm.

However, with the use of a model to incorporate knowledge about human motion, simple corrections can be applied to parts of the silhouette. Of the three factors occlusion is the most challenging to predict and compensate for. The most instances of occlusion take place in the lower half of the silhouette, it is easy to ignore this region for recognition. Most of the upper body is near the centre of rotation, so should be a better choice for orientation independent recognition. Of course a large amount of information is lost with these simple approaches.

Knowledge pertaining to human gait can be applied with increasing sophistication, for example, by only ignoring the lower centre of the silhouette when the legs are close together, preventing thinning and widening of the legs as they are rescaled, and treating the torso as having an elliptical cross section decreases the error in a silhouette's translation or provides more useful data for comparison.

Ignoring parts of the silhouette where the legs cross is unappealing, as the motion of the legs is an important part of human gait. As the orientation becomes further from a normal view the fact that the motion of the legs is not in the same plane as the centre of rotation becomes increasingly apparent. Increasing the complexity of the model and how it is applied allows more of the silhouette data to be intelligently interpreted at any orientation.

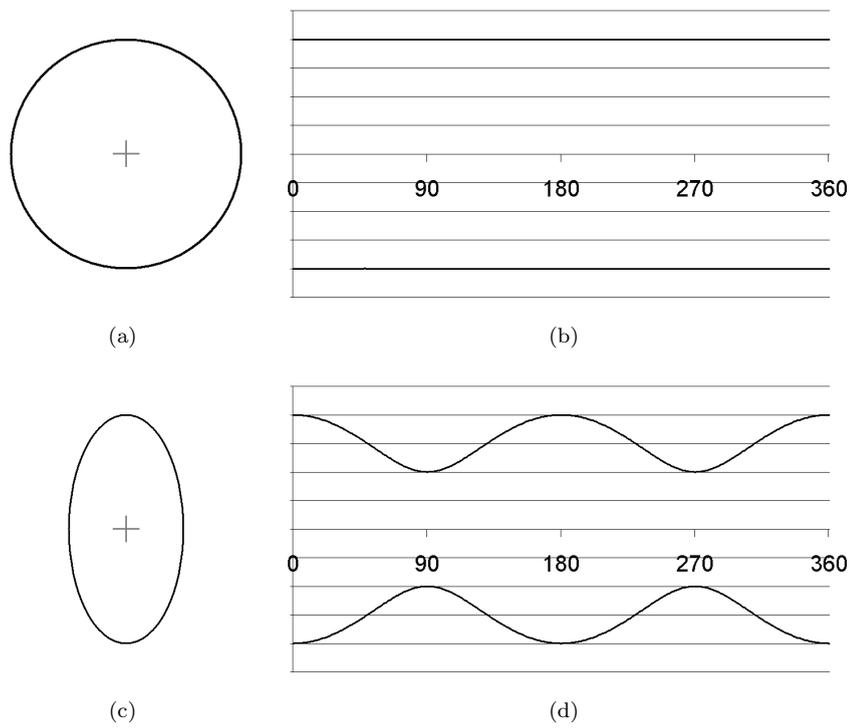


FIGURE 5.8: Two simple shapes (a and c) and outer edges projected onto a plane as they are rotated about their cross (b and d),  $0^\circ$  is horizontal.

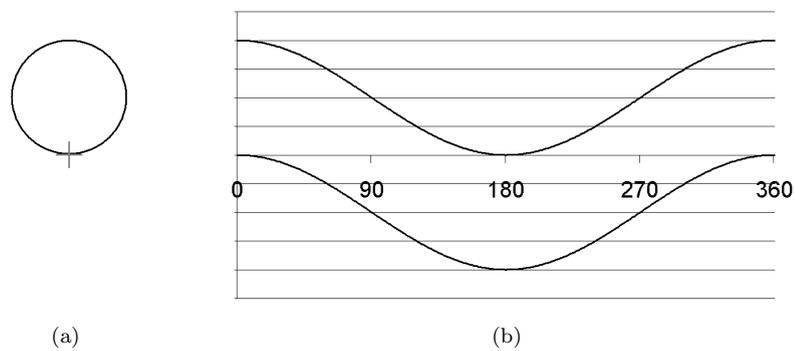


FIGURE 5.9: The projected outer edge positions (b) as the shape (a) is rotated about the cross,  $0^\circ$  is horizontal.

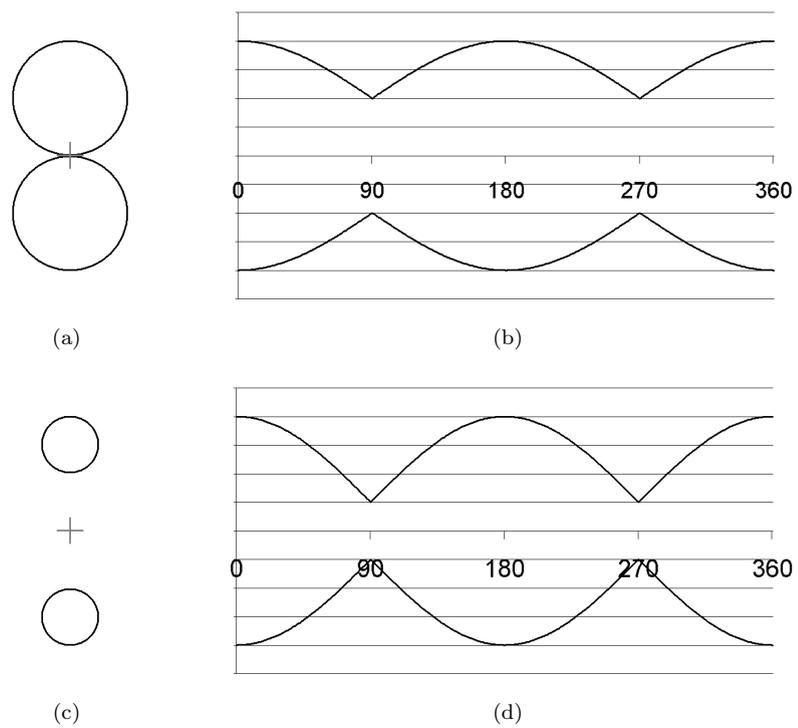


FIGURE 5.10: Two shapes (a and c) and outer edges projected onto a plane as they are rotated about their cross (b and d),  $0^\circ$  is horizontal.

## Chapter 6

# Orientation Independent Model Based Recognition

As discussed in the previous chapter, due to the complex nature of a silhouette projection, occlusion and parallax, it appears unlikely that a 2D transformation can be found to rectify the error introduced by the change in orientation. Given the complex 3D nature of the body, to make sense of a silhouette the use of prior knowledge in the form of a 3D model is essential. Such a model gives the potential to transform the silhouette to a normal view or determine orientation independent parameters describing the 3D model itself.

Calculating the pose of a subject is non-trivial, once the pose of the subject is obtained it may be possible to use a 3D model to normalise the silhouette. One of the simpler methods would be to use a set of affine transforms upon segments of the original silhouette. With an increasingly detailed and accurate model of the subject, more complex an algorithm may be found that generates accurately normalised silhouettes. This may be impractical, require multiple cameras or the estimation of the appearance of unobserved parts of the silhouette. Acquiring the subject's gait accurately enough to enable such a transformation of the silhouette begs the question 'Why not use the gait to recognise the subject?'

Extracting projected limb position based on edges and lines can be unreliable in a gait recognition scenario; the video may be poor quality, low resolution, greyscale with variable lighting and shadows. However, a considerable amount of effort has gone into generating good silhouettes for silhouette based algorithms, basing an algorithm on silhouettes leverages this technology.

One class of 3D pose recovery methods estimate joint locations based upon finding primitives (corners, shapes or patterns), limb length, collision and angular constraints are then propagated through the set of joints found. The final estimated position of

the joints is then used to predict their position in the next frame. A significant amount of information and many constraints are known about the motion and joint positions, however, accurately finding primitives in poor quality video is difficult, if at all possible. Kinematics maps 3D joint angles with a projection to image features. Inverse kinematics attempts to invert this mapping and calculate 3D joint angles from image features, thus requiring accurate and robust feature extraction. Robust feature detection from noisy self-occluding silhouettes is difficult, under these conditions it is unlikely it would be possible to reliably extract the features necessary for inverse kinematics.

The algorithm proposed here extracts the pose of the subject by creating a 3D model and using it to project silhouettes at the same orientation as the subject. By manipulating the parameters describing the 3D pose of the model the difference between the projected silhouette and the real silhouette can be minimised. It is assumed that when the difference is minimal the pose of the model is that of the subject. This can be applied at any orientation and the person specific parameters describing the pose of the model used for recognition. It is hoped that by making use of the whole silhouette rather than attempting to find and track specific points, this algorithm will be more robust than other possible approaches.

## 6.1 The Model

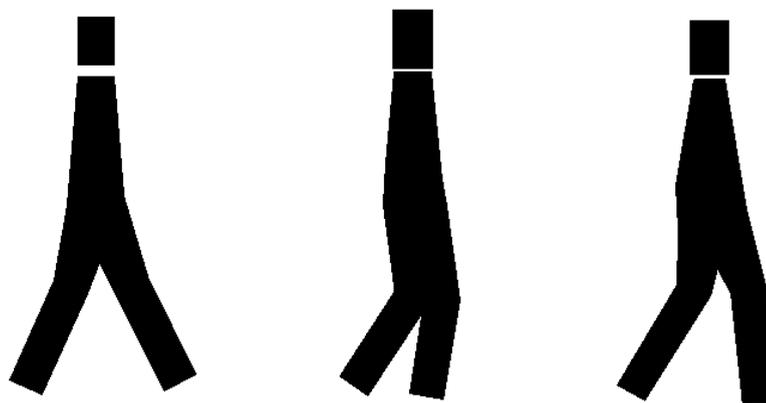


FIGURE 6.1: Three example silhouettes generated from the model.

Figure 6.1 contains three examples of the silhouettes generated from the model. The arms are not modelled as they are of little use for gait recognition; they are relatively small, often occluded and easy disguised when in pockets or carrying something, they would be a significant complication to the optimisation. The feet, being small and complicated, are ignored whilst the problem of accurately aligning the legs is tackled.

The basis of the model is a description of the position of rigid body parts, the shins, thighs, hips, torso, neck and head. To describe the 3D pose of a subject, the length and

Name	Between	Rotates about	Primarily effects
$H_Y^R$	origin and right hip	Y axis	X coordinate of right hip
$H_X^R$	origin and right hip	X axis	Y coordinate of right hip
$H_Y^L$	origin and left hip	Y axis	X coordinate of left hip
$H_X^L$	origin and left hip	X axis	Y coordinate of left hip
$T_Z^R$	right hip and right knee	Z axis	X coordinate of right knee
$T_X^R$	right hip and right knee	X axis	Z coordinate of right knee
$S_Z^R$	right knee and right ankle	Z axis	X coordinate of right shin
$S_X^R$	right knee and right ankle	X axis	Z coordinate of right shin
$T_Z^L$	left hip and left knee	Z axis	X coordinate of left knee
$T_X^L$	left hip and left knee	X axis	Z coordinate of left knee
$S_Z^L$	left knee and left ankle	Z axis	X coordinate of left shin
$S_X^L$	left knee and left ankle	X axis	Z coordinate of left shin
$B_Z$	origin and shoulders	Z axis	X coordinate of shoulders
$B_X$	origin and shoulders	X axis	Z coordinate of shoulders
$N_Z$	shoulders and head	Z axis	X coordinate of head
$N_X$	shoulders and head	X axis	Z coordinate of head

TABLE 6.1: Angles required to fully pose the body in 3D.

orientation of each body segment is described. A subject centred coordinate system is used with the origin at the centre of the hips. For the subject the X axis is forwards, Y is up and Z is to the right, this is a right-handed coordinate system. The notation used here to describe the position of the body is as follows, angle  $a_c^b$  is of body part  $a$  where  $a$  can be  $H$  (hip),  $S$  (shin),  $T$  (thigh),  $B$  (back or body) or  $N$  (neck),  $b$  can be either  $L$  or  $R$  denoting the left or the right side of the body and  $c$  is the axis about which the rotation occurs ( $X$ ,  $Y$  or  $Z$ ). Table 6.1 describes the angles required to pose the lower body. In addition to the angles describing the orientation of the body part, its length is also required to complete its 3D description, the thigh length ( $T_L$ ), shin length ( $S_L$ ), hip width ( $W_L$ ), back (or body) length ( $B_L$ ) and neck length ( $N_L$ ).

There are several steps between the parameters in 6.1 and a silhouette that can be compared to a real silhouette. First the parameters are translated into a set of 3D co-ordinates of joints.

$$\begin{aligned}
\mathbf{rHip} &= \begin{bmatrix} \sin H_Y^R * \cos H_X^R \\ \cos H_Y^R * \sin H_X^R \\ \cos H_Y^R * \cos H_X^R \end{bmatrix} * \frac{H_L}{2} \\
\mathbf{rKnee} &= \mathbf{rHip} + \begin{bmatrix} \sin T_Z^R * \cos T_X^R \\ -\cos T_Z^R * \cos T_X^R \\ \cos T_Z^R * \sin T_X^R \end{bmatrix} * T_L \\
\mathbf{rAnkle} &= \mathbf{rKnee} + \begin{bmatrix} \sin S_Z^R * \cos S_X^R \\ -\cos S_Z^R * \cos S_X^R \\ \cos S_Z^R * \sin S_X^R \end{bmatrix} * S_L \\
\mathbf{shoulders} &= \begin{bmatrix} \sin B_Z * \cos B_X \\ -\cos B_Z * \cos B_X \\ \cos B_Z * \sin B_X \end{bmatrix} * B_L \\
\mathbf{head} &= \mathbf{shoulders} + \begin{bmatrix} \sin N_Z * \cos N_X \\ -\cos N_Z * \cos N_X \\ \cos N_Z * \sin N_X \end{bmatrix} * N_L
\end{aligned}$$

Three more vectors are created representing the left leg, **lHip**, **lKnee** and **lAnkle**. The left side vectors are calculated from  $H_{Y,X}^L, S_{Z,X}^L$  and  $T_{Z,X}^L$  in a similar manner to those of the right.

Not all of the parameters in Table 6.1 are used, the motion of the hips and sideways motion of the legs, torso and neck is ignored, this motion is relatively small compared to the motion of the legs in the X-Y plane. The motion of the left leg is that of the right but  $\pi$  out of phase, to generate a list of the  $T_Z^L(t)$  values using the harmonics describing  $T_Z^R(t)$  in Equation 6.5 we can use:

$$T_Z^L(t) = T_Z^R(t + \frac{1}{2\omega}) \quad (6.1)$$

This simplifies the previous formulae to:

$$\begin{aligned}
\mathbf{rHip} &= \begin{bmatrix} 0 \\ 0 \\ \frac{H_L}{2} \end{bmatrix} \\
\mathbf{rKnee} &= \mathbf{rHip} + \begin{bmatrix} \sin T_Z^R \\ -\cos T_Z^R \\ 0 \end{bmatrix} * T_L \\
\mathbf{rAnkle} &= \mathbf{rKnee} + \begin{bmatrix} \sin S_Z^R \\ -\cos S_Z^R \\ 0 \end{bmatrix} * S_L \\
\mathbf{shoulders} &= \begin{bmatrix} \sin B_Z \\ \cos B_Z \\ 0 \end{bmatrix} * B_L \\
\mathbf{head} &= \mathbf{shoulders} + \begin{bmatrix} \sin N_Z \\ \cos N_Z \\ 0 \end{bmatrix} * N_L
\end{aligned}$$

## 6.2 Silhouette Generation

Given the 3D coordinates of the hips, ankles and knees, the next step is to project them into 2D before constructing a silhouette around them. The matrix for a rotation about the Y axis by  $\beta$  is:

$$R_y(\beta) = \begin{bmatrix} \cos \beta & 0 & -\sin \beta \\ 0 & 1 & 0 \\ \sin \beta & 0 & \cos \beta \end{bmatrix} \quad (6.2)$$

and the parallel projection is:

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad (6.3)$$

The 3D points at orientation  $\theta$  are projected into an image using parallel projection  $P$  and rotation matrix  $R_y\theta$ :

$$\begin{aligned}
\mathbf{rHipIm} &= P * R_y(\theta) * \mathbf{rHip} \\
\mathbf{lHipIm} &= P * R_y(\theta) * \mathbf{lHip} \\
\mathbf{rKneeIm} &= P * R_y(\theta) * \mathbf{rKnee} \\
\mathbf{lKneeIm} &= P * R_y(\theta) * \mathbf{lKnee} \\
\mathbf{rAnkleIm} &= P * R_y(\theta) * \mathbf{rShin} \\
\mathbf{lAnkleIm} &= P * R_y(\theta) * \mathbf{lShin} \\
\mathbf{shouldersIm} &= P * R_y(\theta) * \mathbf{shoulders} \\
\mathbf{headIm} &= P * R_y(\theta) * \mathbf{head}
\end{aligned}$$

Code	Name
$N_W^D$	Head Diameter
$N_W^H$	Head Height
$B_W^T$	Body Width at the Top
$B_W^B$	Body Width at the Bottom
$T_W^T$	Thigh Width at the Top
$T_W^B$	Thigh Width at the Bottom
$S_W^B$	Shin Width at the Bottom

TABLE 6.2: Parameters used to draw a silhouette.

Once the 2D co-ordinates are calculated ( $\mathbf{rHipIm} \dots \mathbf{headIm}$ ) a silhouette can be drawn about them. The lower legs, thighs and torso are assumed to be cylindrical or conical, their projections drawn as rectangles and trapezoids. This is inaccurate at the ends of the cylinders but computationally simple. Unless the cylinder's major axis is parallel to the image plane (not tilted towards or away from the camera) the end of the shape should appear curved, however, only three ends are visible; the top of the torso and at the ankles. The torso remains approximately vertical and as such the simple projection will not be significantly different to the correct projection. The discrepancy between a rectangle and a projected cylinder will be much less than that between either and a real foot.

The appearance of the shins, thighs and torso is parameterised with an additional two parameters per body part that describe the thickness of the body part at each end. Both shins and thighs appear the same and share parameters at the knees. The model defines the position of the centre of the head but not its appearance, it models it as a vertical cylinder, requiring one parameter for the diameter and one for the height. Table 6.2 lists these parameters and Figure 6.2 shows how all the parameters are used to draw silhouettes such as the three in Figure 6.1.

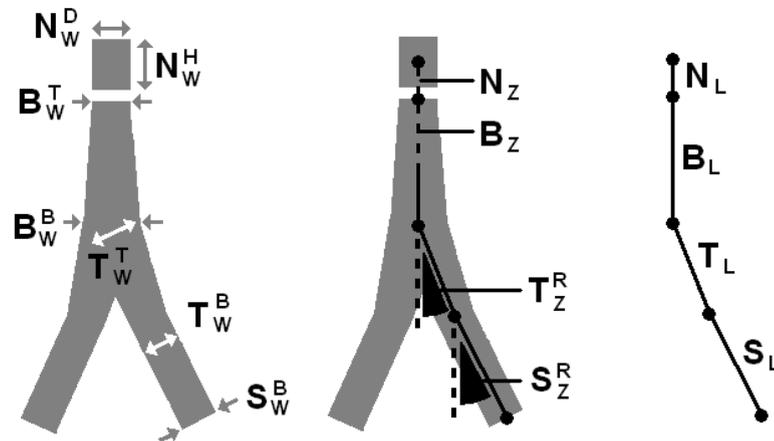


FIGURE 6.2: The parameters used to generate a silhouette.

### 6.3 Silhouette Alignment Optimisation

To ultimately infer real life leg positions the generated silhouette must be aligned with the real life silhouettes. No effort is made to model the true 3D position of the subject, leg angles are sufficient. The generated silhouette is aligned with the real life silhouette at the hips. The position of the centre of the hips in our generated image is known, the centre in the real image is estimated using anthropometric data based upon the silhouette's height, in the horizontal centre of the silhouette.

The estimation of the hip position in a real image may be inaccurate so we use additional parameters for the x and y offsets ( $H_X^O, H_Y^O$ ) to allow adjustment of the hip position relative to the silhouette with the other parameters. The hip width  $H_L$  is not optimised, an anthropometric mean is used instead; during preliminary tests the hip width could not be reliably aligned to that of the silhouette sequence, this is probably due to it being more difficult to observe than leg lengths or angles. Also its value (which was inaccurate) has a large negative impact on the resulting alignment of other parameters.

The model now consists of 19 parameters describing the silhouette that represents a single frame that can be adjusted to change the generated silhouettes to match the real silhouettes:

- $N_Z, B_Z, T_Z^R, T_Z^L, S_Z^R, S_Z^L$ . The angles defining the position of the body parts.
- $N_L, B_L, T_L, S_L$ . The lengths of the body parts, with the previous angle they are enough to draw stick figure.
- $N_W^H, N_W^D, B_W^T, B_W^B, T_W^T, T_W^B, S_W^B$ . The widths of the of the body parts.
- $H_X^O, H_Y^O$ . x,y offsets for the model.

Fitting the artificial silhouette to the real silhouette is an optimisation problem, with the goal of finding the parameters that minimise the distance between the real and artificial silhouettes.

The distance between two silhouettes is calculated as the number of different pixels between two ( $S^1, S^2$ ) silhouettes:

$$\begin{aligned} D(S^1, S^2) &= \sum_{x=0}^X \sum_{y=0}^Y (S_{x,y}^1 | S_{x,y}^2) - (S_{x,y}^1 \& S_{x,y}^2) \\ &= \sum_{x=0}^X \sum_{y=0}^Y (S_{x,y}^1 \odot S_{x,y}^2) \end{aligned}$$

where  $\odot$  is the XOR operator. Summing  $D$  over a  $N$  long sequence, the total number of different pixels between two silhouette sequences is:

$$Dist(a, b) = \sum_{n=0}^N D(R^n, A^n) \quad (6.4)$$

Figure 6.3 shows the legs (where most of the difference is) of two silhouettes and their overlap, the black area of Figure 6.3(c) are the pixels being counted.

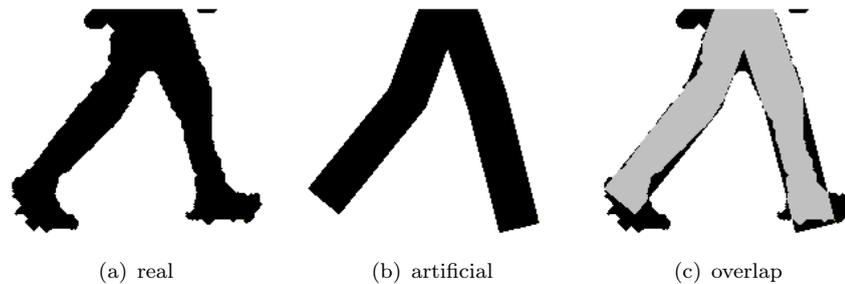


FIGURE 6.3: Real (a) and artificial (b) silhouettes with their overlap (c), white is background, grey is the overlap, black is the difference between the two.

The motion parameters are initialised with mean motion of the motion capture data (Geisheimer, 2001) used to animate the SGAW Section 3.1. The leg lengths and widths are initialised with mean anthropometric values based upon the height of the silhouette.

Five different optimisation algorithms were assessed for fitting the artificial silhouettes to the real data, first was a very simple local search algorithm:

```
while change > minChange:
    for each parameter:
        calculate Dist(R,A) with this parameter
        calculate Dist(R,A) with this parameter + change
        calculate Dist(R,A) with this parameter - change
        change this parameter to the value with the lowest Dist(R,A)
    if no parameters were changed:
        change /= 2
```

this algorithm is referred to as a local search because its inefficiencies will be pronounced if not initialised near the optimal value.

The axes aligned linear minimizations from Press et al. (1992) implemented and is potentially faster. Brent's method for linear minimisations applied to each axis of the 19 axes in a round robin fashion until a minimum is found, Brent's method is able to travel large distances but is limited to the directions represented by the axes.

Powell's direction set method aims to alleviate the limitations of travelling only in the directions of each axis by modifying the set of directions to travel in. The set of directions is updated with 'good' directions as they are found.

The Nelder-Mead simplex algorithm maintains a set of points and iteratively replaces the worst point by projecting it through the remaining points.

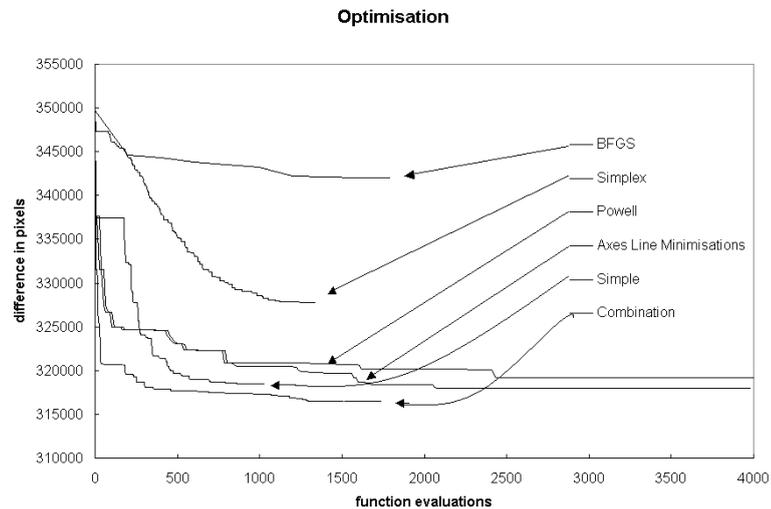


FIGURE 6.4: The speed and depth of the several different optimization techniques.

The Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm is a quasi-newton method that assumes the landscape near the minimum is approximately quadratic. The hessian matrix of second derivatives is used but not calculated, rather it is updated at each iteration.

Powell's direction set method, the Nelder-Mead simplex algorithm (Nelder and Mead, 1964) and the BFGS method were implemented from Press et al. (1992).

The results shown in Figure 6.4 indicate that the simple algorithm converges to the best solution fastest, however it is relatively slow starting. Performing an initial line search along each axes before commencing the simple search speeds up the optimisation, this is the combination minimisation of Figure 6.4. The 19 dimensional landscape that these algorithms are attempting to minimise over is going to be extremely complex, a downhill path from an arbitrary point following a 'twisty' path to an optima. A gradient based method will have many problems in this situation (i.e. BFGS), the simplex method appeared to get some vertices stuck in separate valleys. The best method is the simple local search, it behaves in a similar way to simulated annealing with change as temperature but ignores gradient.

Figure 6.5 shows the extracted motion of the thigh and shin that results from the minimisation of the difference and subsequent extraction of limb angles from the model. It can be seen clearly that there are errors in this estimation, spikes in the traces that would indicate very fast stuttering back and forth of the thigh and shin.

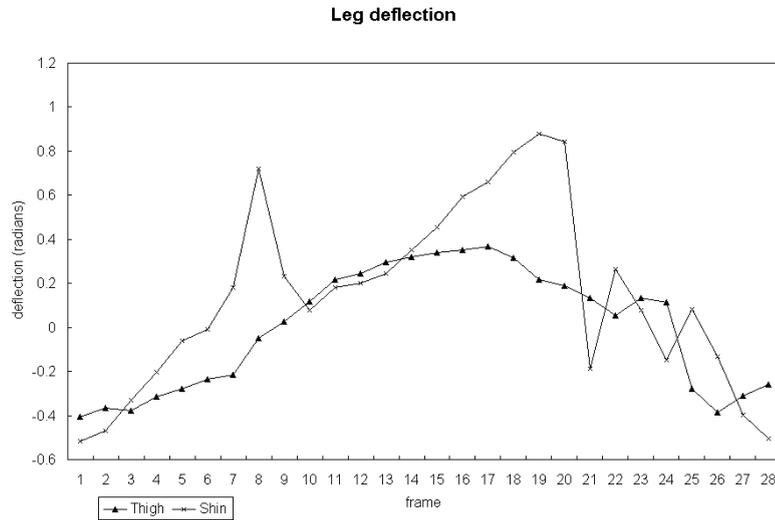


FIGURE 6.5: Extracted thigh and shin motion.

## 6.4 Improved initialisation

Each optimisation of each frame is independent of the others, i.e. there is nothing enforcing consistent leg lengths or motion between frames of the video sequence. In practice this leads to ‘spiky’ motion (see Figure 6.5) where optimisation finds a false minimum, the generated silhouette appears similar to the real silhouette but is constructed incorrectly. One option is to impose inter-frame constraints upon the parameters, enforcing smoothness of angle motion or consistent leg lengths, this complicates and slows optimisation.

Any optimisation method can benefit from an improved initialisation and the current optimisation was initialised from the SGAW with its peculiar limb lengths and angles. The initialisation involved calculating the average motion of all the SGAW data, after it had been aligned and normalised. A Fourier series was fitted to the data and this subsequently used to provide initial values for the optimisation with real data. In [Cunado \(1999\)](#) and [Carter and Nixon \(2000a\)](#) the components of a Fourier series were used for recognition and it was found that only the first 7 harmonics were required to describe human gait, higher frequency harmonics represented noise.

For example, the forwards rotation of the right thigh at time  $t$  can be calculated from a sequence of 7 harmonics of the form  $x_0, x_1 + y_1i, x_2 + y_2i, \dots, x_7 + y_7i$  by

$$T_Z^R(t) = \frac{x_0}{2} + \sum_{n=1}^{\infty} (x_n \cos n\omega t + y_n \sin n\omega t) \quad (6.5)$$

where  $\omega$  is the frequency of the subject’s gait in frames. This can be improved upon to lessen the effect of the SGAW and initialise the optimisation closer to the true minimum

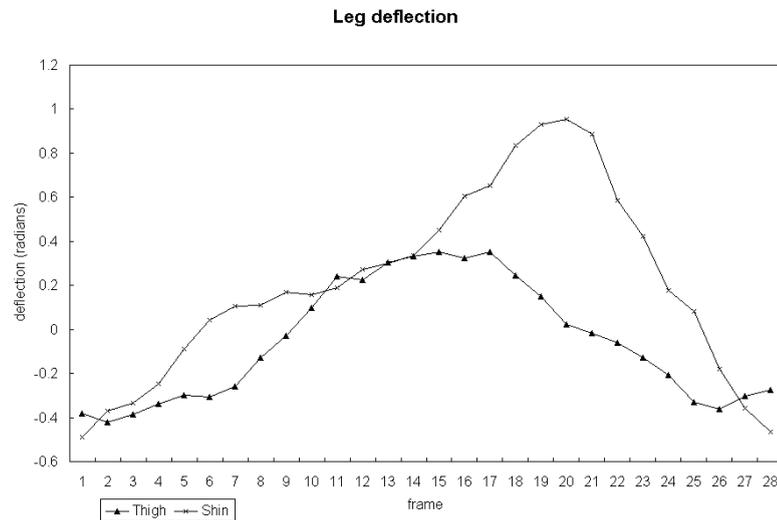


FIGURE 6.6: Extracted thigh and shin motion with improved initialisation.

avoiding any false minimums. To do this each harmonic was adjusted in order using the simple local search above but evaluated over the entire sequence:

```

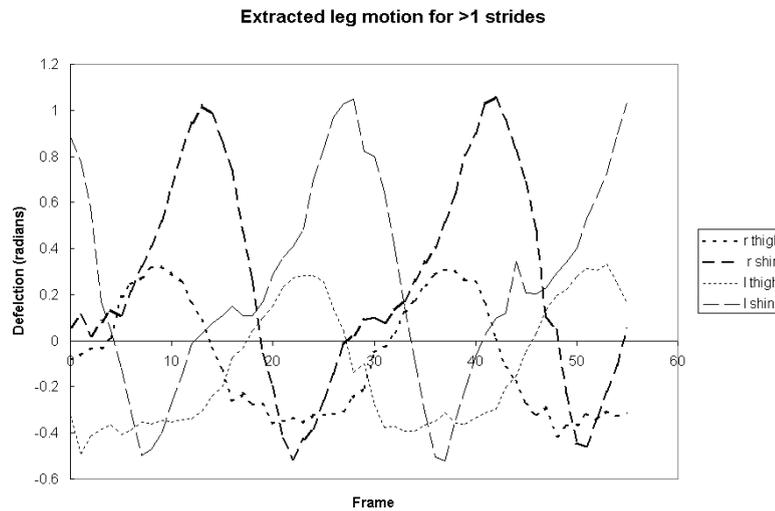
for Harmonic in [0,1,2,3,4,5,6,7]
    change = 10\% of the magnitude
while change > minChange:
    for x then y:
        calculate difference over the whole sequence with this parameter
        calculate difference over the whole sequence with this parameter + change
        calculate difference over the whole sequence with this parameter - change
        change this parameter to the value with the lowest Dist(R,A)
    if no parameters were changed:
        change /= 2

```

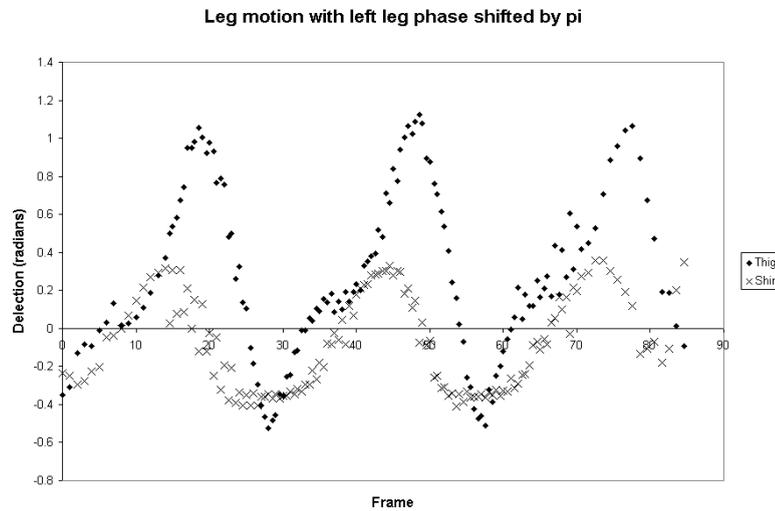
Starting with the lowest frequency components each is manipulated in turn until the alignment can not be improved. This method of optimisation is slow, the difference over the entire sequence is evaluated at each point not each frame. Manipulating the angles describing the models' motion via a Fourier series leads to a poor final alignment, however a single iteration makes a large improvement to the initialisation, avoiding local minima Figure 6.6.

## 6.5 Recognition

The widths and lengths of the body parts are separate for each frame and potentially inconsistent, they are not used for recognition as they do not describe a persons gait.



(a)



(b)

FIGURE 6.7: Motion extracted from more than one stride (a), and phase shifted by pi (b).

To generate features the Fourier series that best fits the extracted angles is used for comparison. To compare two Fourier series the sum of the differences in magnitude is used for K-NN. Extracting parameters describing a sequence longer than a single stride is simple. Figure 6.6 shows the angles extracted from a series of frames (Figure 6.7(a)) wrapped to one period and the resulting motion estimate (Figure 6.8(a),(d)). Once a set of Fourier series (one for the thigh, one for the shin) describing the motion has been calculated, it can be compared to other series with the phase-weighted magnitude method of Cunado (1999).

The model was applied to the Southampton gait database and gait signatures generated in the form of Fourier series. From the Southampton database 8 sequences of 50 people

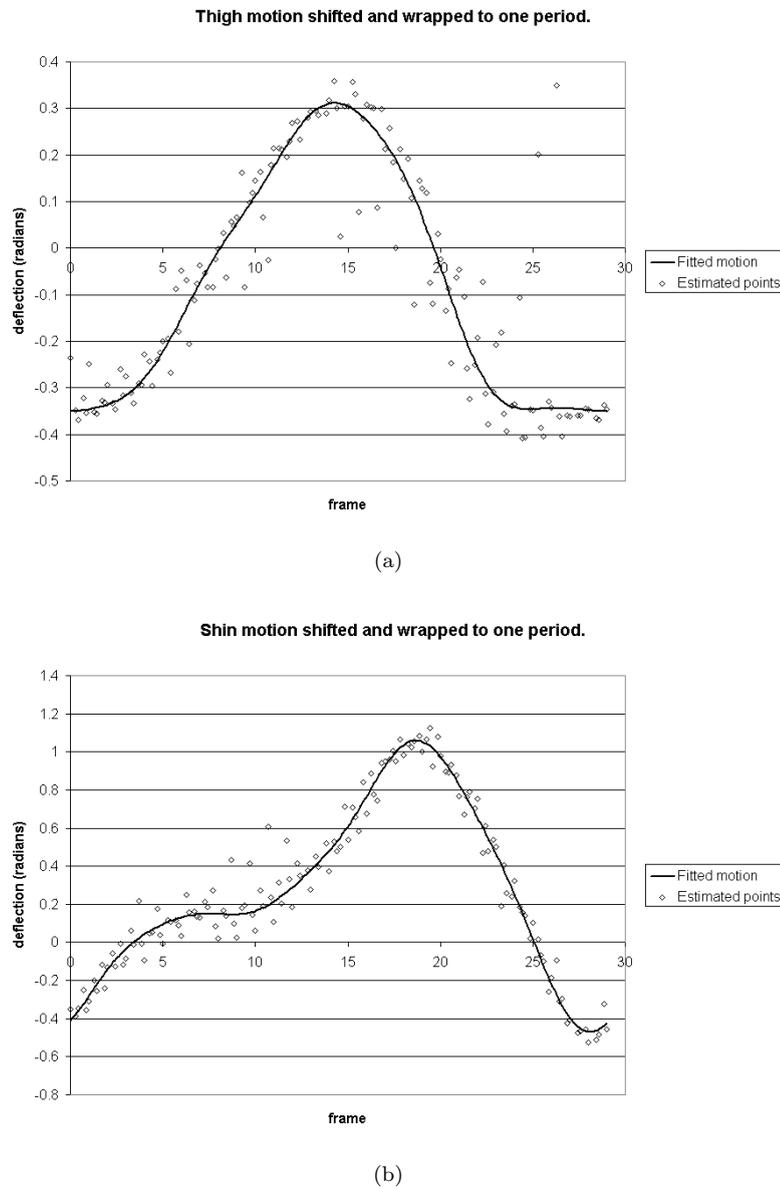


FIGURE 6.8: Motion extracted from more than one stride (a), left leg shifted by  $\pi$ , and wrapped to one period (b).

were used, two instances, walking left and right at two orientations, 0 (normal) and 22 degrees from normal. The silhouettes were cropped around their centroids to align them. A leave one out cross validation recognition experiment was performed, using one of the 4 sequences as the probe and the other 3 of all subjects as the gallery. This is an identical setup to the average silhouette experiment in Section 4.5

A round robin leave one out cross validation experiment was performed on the real silhouettes identical to the one carried out in Section 4.5. Two sequences of the subject walking left then right from two cameras of fifty people were used. Table 6.3 shows the results of the recognition experiment.

	gallery	
	$0^\circ$	$22^\circ$
probe	$0^\circ$	$10\%$
	$22^\circ$	$16\%$

TABLE 6.3: Recognition rates with the model based algorithm.

## 6.6 Discussion and Conclusions

Although the absolute recognition rate in Table 6.3 is not as good as that achieved in Section 4.5, the relative drop in performance between cameras is much less.

There is much room for improvement and several places improvements can be made. Yam et al. (2001), Cunado (1999) and Carter and Nixon (2000a) used the motion of the legs to perform recognition and results competitive with the average image algorithm. Therefore, the poor performance is likely due to inaccurate estimates of the leg motion. The Traces in and most noticeably contain anomalies and large amounts of noise. Further improving the accuracy of the alignment will result in more accurate angles extracted per frame, this will improve recognition.

The model used here may not be able to accurately match the shape of the silhouettes, the positions found may be the globally optimal configurations but unable to select the real world pose of the subject. A more complex model that allows us to generate more accurate silhouettes would reduce this problem at the expense of complexity and computational cost of the optimisation. The manner the 2D image is generated places disproportionate emphasis on some model parameters. For example, the position of the centre of the hips effects the angle and length of the thigh which effects the shins. Unfortunately the centre of the hips is not directly observable in the silhouette. Lowering the importance such parameters or the effect of inaccurately estimating them would relax hindering constraints on other parts of the model.

The distance metric that is measured and minimised by the optimisation is insensitive to the task of aligning human silhouettes, differences at any pixel are counted equally. Pixels that are different due to noise at the edge of the silhouette are less important than different pixels far from the subject.

In conclusion, though capable of extracting approximate motion of the subject the algorithm is incapable sufficient accuracy for reliable recognition. The low recognition rate is still encouraging as there are many ways of improving the process.

## Chapter 7

# Improved Model Based Algorithm

The recognition rates recorded in the previous chapter are encouraging due to the relatively small drop in performance between view compared to recognition within a single view. However the absolute recognition rate was not as high as that achieved by the average silhouette algorithm.

We can ask the question, ‘What is the root of the poor recognition rate?’ Figure 6.6 shows wildly inconsistent angle estimation during a single sequence. This is due to the poor alignment of the artificial silhouette with the real one. The optimisation has found a false minima or, more worryingly, finding the global minima but the global minima corresponded to a pose that does not accurately reflect the true pose of the subject. This problem was alleviated with improved initialisation, however it is symptomatic of an ‘unfriendly’ optimization landscape, a landscape requiring a convoluted path from an initialization point to the global minima, containing convincing (in terms of the difference function) false minima. In this chapter modifications of the algorithm are explored with the aim of providing a ‘friendlier’ landscape to traverse to the global optimum.

### 7.1 Improved Fitness Metric

The fitness function is currently the size of the XOR of the synthetic images and the real silhouettes from the sequence, this has the advantage that it is computationally and conceptually simple, however it also has the disadvantage that all parts of the synthetic silhouette that do not overlap the real silhouette count equally towards the fitness (or unfitness) irrespective of their distance from the real silhouette. A single pixel difference adjacent to the real silhouette contributes to the difference by the same degree as a pixel several body widths away. This method of measurement is counter intuitive, a differing pixel adjacent to a body part may be due to noise in the original silhouette, and is more tolerable than a differing pixel far from the body.

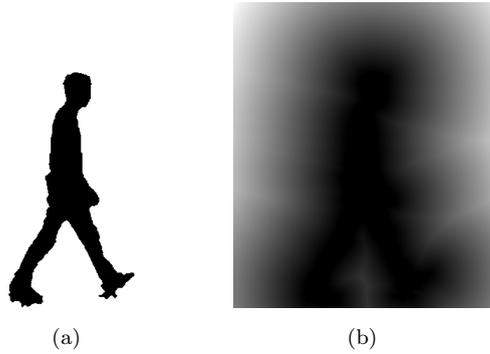


FIGURE 7.1: The real silhouette used in comparison (a) and the corresponding distance map (b).

This is addressed by pre-processing the real silhouettes, first finding the edges then calculating the Euclidean distance map. The Euclidean distance map contains at each pixel the distance to the nearest part of the silhouette. Figure 7.1

The fitness function is modified so that the XOR of the real ( $R$ ) and artificial ( $A$ ) silhouettes is multiplied by the distance map ( $M(R)$ ) of the real silhouette. With the size the silhouette frames being  $(X, Y)$ :

$$D(A, R) = \sum_{x=0}^X \sum_{y=0}^Y (A_{x,y} \odot R_{x,y}) * M(R)_{x,y} \quad (7.1)$$

where  $\odot$  is the XOR operator. The per pixel difference is now greater for pixels further from the real silhouette. The effect this will have upon the landscape is that directions used in optimization that trade near body pixels for far from body pixels have an increasing difference instead of a constant one.

## 7.2 Using head ankle angle

The thigh angle was found to be difficult to extract but when measured accurately a powerful biometric. The main hindrance to accurately aligning the model thigh with the silhouette is that the thighs commonly occlude each other. Additionally, the shape of the top of a leg is often occluded by baggy trousers, torso clothing that descends lower than the hips and the hands as they swing past.

These problems are sidestepped by discarding the thigh angle biometric in favour of the head to ankle angle shown in Figure 7.2. The head-ankle angles are denoted as  $A_Z^R$  and  $A_Z^L$  for the right and left ankles respectively, the distances are  $A_L^R$  and  $A_L^L$ . Unlike other lengths in the model there is a left and right length, this is because unlike others it is expected to vary over time and between legs. These parameters are not used during the

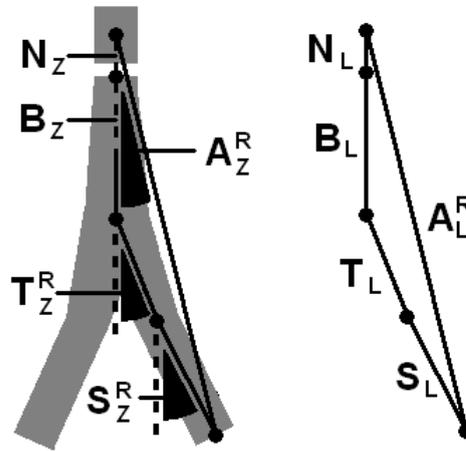


FIGURE 7.2: The new head ankle measurements in relation to the existing parameters.

optimization to align the model, they are calculated from the final aligned position of the head and ankles.

These head-ankle parameters are a function of the others, including the thigh and shin. The advantage is that even if the thigh (or other parameters) are aligned incorrectly, as long as the head and ankle are accurate the new parameters should be accurate. The new parameters are not independent of the shin parameters, but neither are those of the thigh so the reduction in the orthogonality of the set of parameters (head-ankle and shin from thigh and shin) may not be that great. Additionally the motion and length of the back and neck is incorporated so it may be less correlated to shin motion than the thigh.

### 7.3 Removing hip constraint

It was also observed that the position of the hips did not alter significantly from its initialised value. From inspecting the model we can see that changing the position of the hips requires a corresponding change to many parameters to keep the shoulders and knees in the same position. The interdependence of these parameters causes the ripples and false minimum that make the optimisation hard. To demonstrate this we constructed the following example. Two fixed width artificial thighs were aligned, one fixed vertical, the other displaced to the left by  $x$  pixels, its length and angle space were exhaustively searched to find the parameters that minimise the difference between them, shown in Figure 7.3. The optimal length and angle were found for a range of values of  $x$  from -30 to 0. Figure 7.4 and Figure 7.5 show the optimal angle and length change in a complicated manner as  $x$  approaches the global optimal.

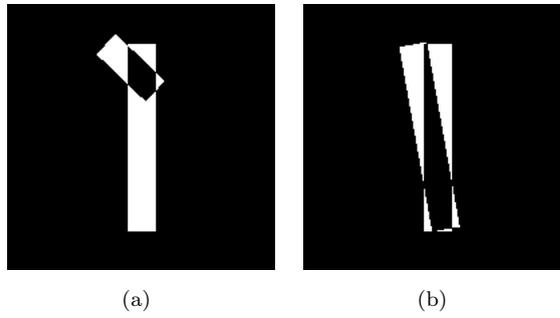


FIGURE 7.3: Two images of aligned synthetic thighs, the x coordinate of the top is fixed, the length and deflection from vertical of the left rectangle is optimised with a brute force algorithm to minimise the difference.

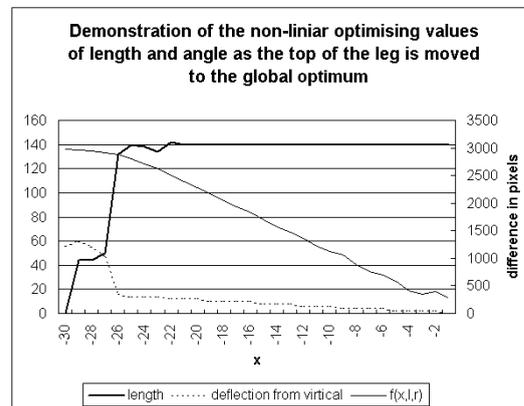


FIGURE 7.4: The fitness and optimising values of angle and length at fixed horizontal displacements.

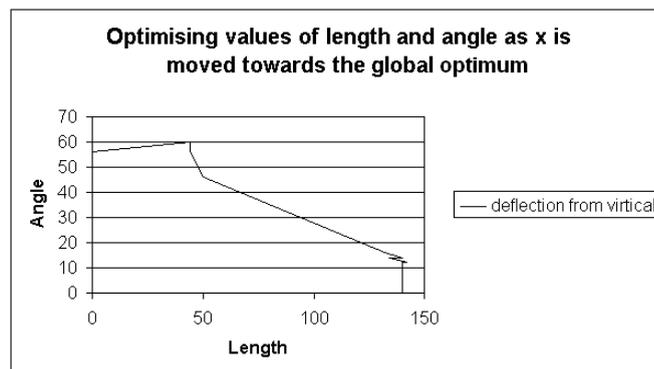


FIGURE 7.5: The optimising values of angle and thigh plotted against each other under various horizontal offsets.

To avoid this complexity the connection at the hips between the legs and the lower back is broken. This may have a negative impact upon the resulting estimation of the thigh motion, however presently the motion of the thigh is not used directly.

## 7.4 Template based lower legs

The model's lower legs are a poor approximation of the real silhouette's, as their primary function is to simulate the lower leg and not the foot. The foot, even though it is a relatively small body part is a source of considerable error. An optimisation algorithm will attempt to align the lower leg with the foot to some extent, this is expected to cause undesirable minimum fitness landscape near the global optima.

The model could be extended to model and draw a foot, this would require additional parameters, two for the orientation in 3D, one for the length and two for the width (at the heel and toes). This would allow the model to place the lower leg at the correct orientation but has some significant drawbacks. The extended model has a correspondingly increased dimensionality fitness landscape that would need traversing to the new global optimum. This complicates the fitness landscape as the parameters describing the foot will be intimately linked to the position of the ankle. A misaligned foot may cause the ankle to be misaligned and consequently the knee, correcting the error would require simultaneous modification of features describing all 3 body parts in a complex manner, 2 optimization iterations steps may read:

1. reduce length of foot and ankle width while rotating clockwise, rotate shin anti-clockwise, thigh clockwise
2. increase ankle width, rotate shin clockwise and increase length of thigh

By adding parameters and consequently dimensions to the landscape the optimisation traverses, the problems mentioned previously that applied to the hip will be encountered to a greater extent. The landscape becomes bumpier and consequently increases false minima near the global optima. Even with the extended model we may be unable to synthesise silhouettes that closely resemble the real feet as this area is extremely noisy; clothing shadows and even laces change the appearance of the feet.

Another method of modelling the feet is proposed here; create template silhouettes of the lower legs and using these instead of the boxes used previously. A 50 by 100 pixel template can be considered to be a list of 5000 parameters, however, once the template is generated these parameters are fixed and need not be part of the optimisation.

Figure 7.6 describes the process we use to generate templates of the lower leg, using the results of the box based fit from the previous chapter the lower legs are segmented out

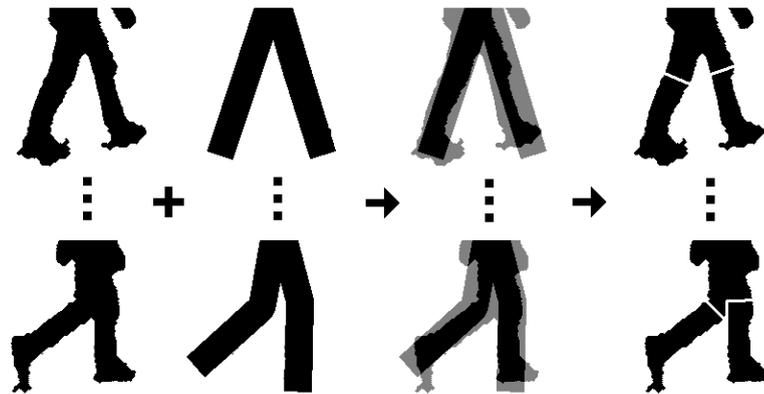


FIGURE 7.6: The first process is the generation of a template based model. Lower legs are segmented out of the original silhouettes.

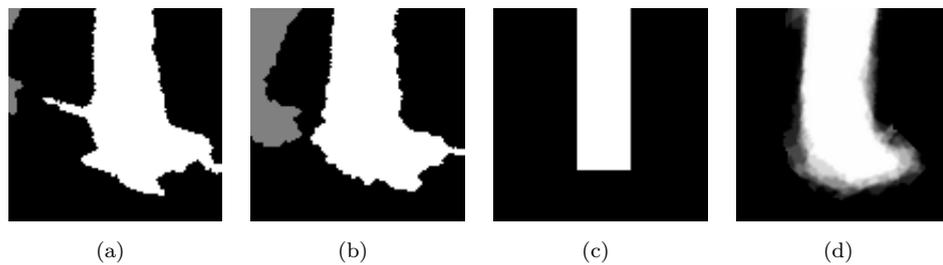


FIGURE 7.7: (a) and (b) segmented legs, the initial template (c) and the final template (d).

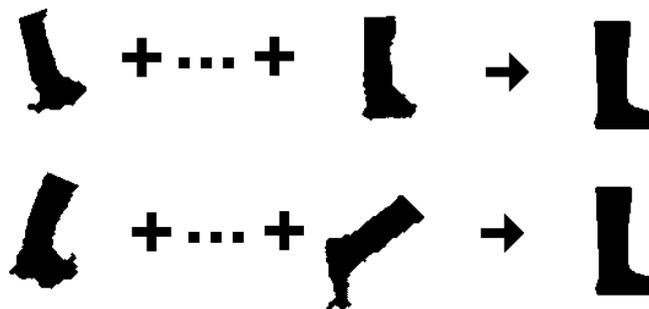


FIGURE 7.8: The second process in generating a template based model. Segmented legs are used to construct an average lower leg.

of the silhouettes. From the existing alignment frames can be selected where the ankles are sufficiently separated so that there is no occlusion between them. Figure 7.7 (a) and (b) show some of extracted legs, based upon the estimates of the shin angle and length the legs are rotated to vertical and translated so the ankles are aligned. Parts of the other leg (shown in grey) are visible and easily removed from the image.

Once the legs have been segmented out of the silhouettes templates are created, Figure 7.8. An initial best estimate of the appearance of the lower leg is created, using that

of a trapezoid (Figure 7.7(c)) constructed with the leg widths previously found. Each of the extracted legs is aligned with its respective trapezoid, the ankle x,y coordinates and the shin deflection from vertical is optimized. The XOR difference function with the simple optimisation algorithm is sufficient. Each aligned leg silhouette is then added to an accumulator for that respective leg. Once all viable frames have been processed a new best estimate is set to the image in the accumulator (Figure 7.7(c)). The process is then repeated, each segmented leg returned to its initial values and re-aligned with the new best estimate of its appearance. This process need only be applied three times and often only twice before there is no further change in the best estimate. Once the template has been created the ankle has to be found.

The point referred to throughout this thesis is actually a point on the sole of the foot that intersects an imaginary line down through the centre of the shin template. This point corresponds to the centre of the bottom of a trapezoid previously used to model the lower leg. This point is easy to find in Figure 7.7(c) before the template is created but may not be in the same position in the final template. Additionally, the final template may not be aligned vertically. This is fixed by performing a last alignment of the left and right leg templates with the trapezoid initialisation.



FIGURE 7.9: A head template created similarly to the leg templates.

Once the templates have been created they are incorporated into the drawing of the artificial silhouettes. A template was created for the head of a silhouette (Figure 7.9), the method is the same as for that of the legs. Having a template for the head increases the consistency of the estimation of the head which is more important with the use of the head-ankle angle.

Replacing the relevant trapezoids with the templates results in silhouettes such as those in Figure 7.10. There are a few minor problems that need to be addressed before the silhouettes are complete.

The templates contain some noise, this is a result of the noise in the original silhouettes, combined with inaccuracies in the segmentation. By thresholding the silhouettes at 50% the majority of the noise is eliminated. The thighs are not connected to the shins, however finding the top of the thresholded thigh and joining them to the thigh trapezoids is simple.

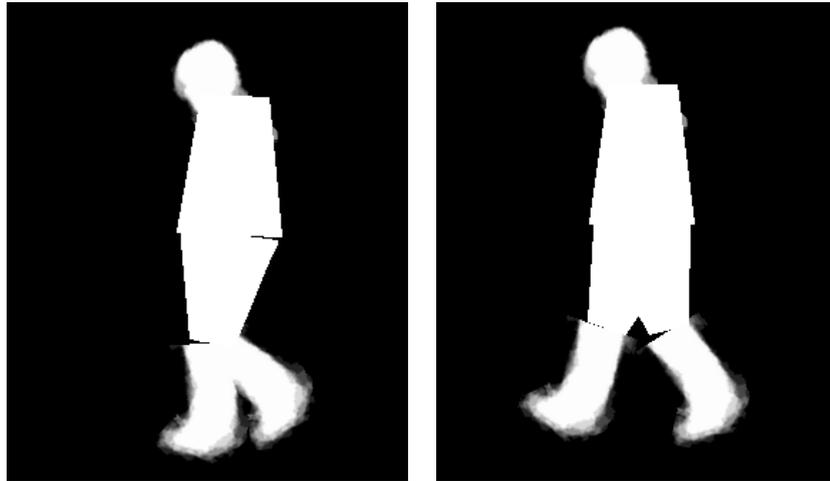


FIGURE 7.10: 2 example silhouettes generated from the template model.

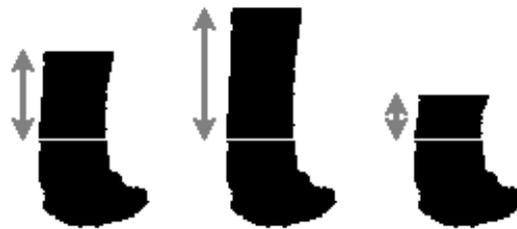


FIGURE 7.11: Rescaling the shin for different shin lengths.



FIGURE 7.12: 2 example silhouettes generated from the template model.

The length of the shins is a parameter dictated by the model, thus a method of modifying the length of the template is required. Scaling the template vertically would change the shape of the foot, instead only the top half of the template is rescaled, Figure 7.11 shows how this appears. Figure 7.12 shows two example silhouettes generated by the algorithm, in addition to their more realistic appearance they eliminate the need for several parameters,  $N_W^D$ ,  $N_W^H$  and  $S_W^B$ .

	gallery	
	$0^\circ$	$22^\circ$
probe	$0^\circ$	$28\%$
	$22^\circ$	$30\%$

TABLE 7.1: Recognition rates using the new template fitting algorithm between orientations.

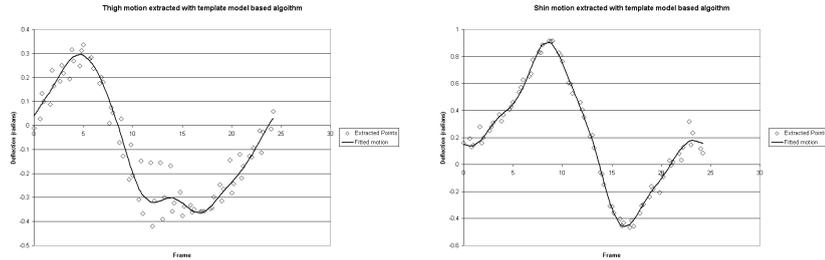


FIGURE 7.13: 2 example silhouettes generated from the template model.

## 7.5 Results

In summary, the lower legs and head have been replaced with templates generated from the sequence itself; the constraint holding the back and thighs together at the hips is broken, the thigh angle information has been discarded in favour of the heel to head angle and implemented a distance based fitness function to drive the optimisation towards the global optima.

The improved model was applied to the Southampton GaitDatabase and gait signatures generated in the form of Fourier series. From the database 8 sequences of 50 people were used, two instances, walking left and right at two orientations, 0 (normal) and 22 degrees from normal. The silhouettes were cropped around their centroids to align them. A round robin leave one out cross validation recognition experiment was performed, using one of the 4 sequences as the probe and the other 3 of all subjects as the gallery. This is an identical setup to the average silhouette experiment in Section 4.5 and Section 6.5.

Figure 7.13 shows the thigh and shin motion calculated from an image sequence, the shin motion especially is much more consistent than that recorded with in the previous chapter (Figure 6.6). The recognition rates achieved (Table 7.1) are also an improvement upon the previous chapter and show a large degree of tolerance to orientation. .

## 7.6 Discussion and Conclusions

The experiment achieved the recognition rates in Table 7.1. This is a large improvement upon the model based Table 6.3. The relative drop in performance between views is much less than for the simple model based algorithm and average silhouette algorithm

(Table 4.4). Indeed, the between orientation recognition rates are comparable to those achieved with the average image algorithm. In the ideal circumstances, a normal probe with a normal gallery, the recognition rate still has a large potential for improvement, improving this should also improve cross orientation recognition.

The extracted shin angles appear very consistent across however the extracted motion of the thigh is still noisy, and has potential for improvement.

A peculiarity of the results is the significant drop in our recognition ability at the 22° camera. Going back to the original RGB data captured from the cameras, it was discovered that the 22° camera was used in an interlaced mode, the background subtraction was then applied directly to the interlaced images and a binary closing performed to eliminate the feathering. Figure 7.14 shows the problems this causes, body parts are blurred in the direction they are moving proportional to the speed at which they move. The lower legs move in different directions at different parts of the gait cycle with a changing speed.

Figure 7.15 shows the different shape of feet during the sequence, this has a huge detrimental impact upon the template model. Poor quality templates are created as the shape of the foot changes through the sequence, additionally, the width of the shin is different depending on its speed. The silhouettes themselves, with which we aim to align the model, are corrupted, forcing the algorithm to attempt to align a malformed template with a single leg that in some circumstances has two feet. Despite these problems the algorithm still has the capability to recognise subjects. This is a testament to the robustness of gait and the algorithm.

Recognition between orientations has been performed with comparable error to the average image algorithm. However, this algorithm is using only the motion of the legs, this includes the noisy motion of the thighs. If the motion of the thigh can be found more accurately and robustly the base performance and inter-view performance should increase. Another avenue to better performance is to include other observable features such as the height or body shape although these measurements are not part of human gait.

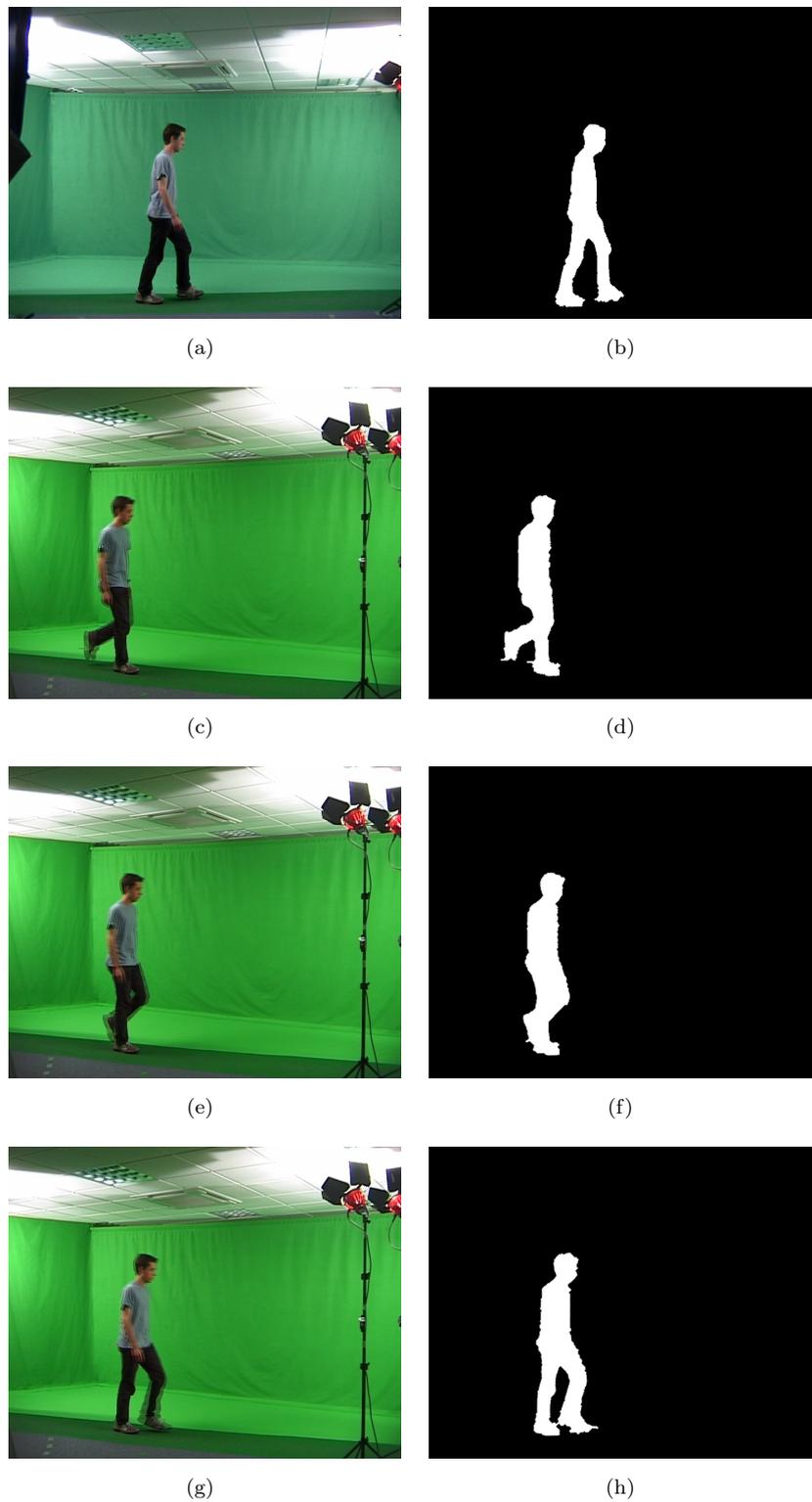


FIGURE 7.14: Captured images and the silhouettes calculated. (a) and (b) an image captured from the  $0^\circ$ . (c) and (d), interlaced image with noticeable vertical blur to the lower leg, (e) and (f), blurring increasing the occlusion of the legs, (g) and (h) horizontal blurring of the lower leg.

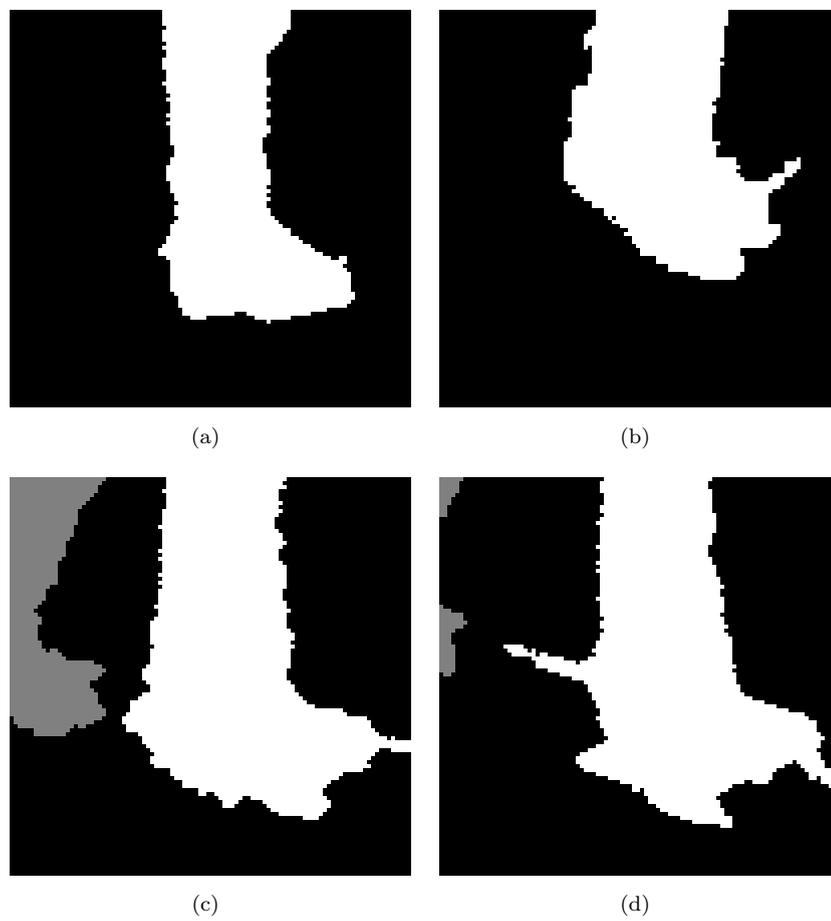


FIGURE 7.15: examples of the same foot segmented from different parts of an interlaced sequence. (a) is stationary with minimal corruption, (b) has vertical blur, (c) and (d) have horizontal blur.

## Chapter 8

# Future Work and Conclusions

### 8.1 Conclusions.

The recognition ability of the baseline algorithm (Chapter 3) was found to degrade quickly when comparing video footage captured at orientations that differed by greater than 15 degrees. This experiment was performed upon artificial silhouettes drawn using motion capture data (Southampton GTRI Artificial Walker), this allowed any view to be synthesized. This is extremely limiting for many gait applications as many views of subjects need to be captured and in some scenarios this is impossible.

The average Image (Chapter 4) was found to have a similar tolerance for orientation. A simplistic correction for differing orientation was attempted found to be of little use at most orientations. The correction involved assuming people are 2D in the plain their legs swing in and rescaling the silhouette accordingly. The results with the SGAW were corroborated with the drop in recognition ability observed real life data.

A moment based algorithm (Chapter 5) was applied to the SGAW data, the hypothesis being that a subset of features generated from from silhouette sequences would have a greater degree of independence than the average image and baseline algorithms. Analysis of moments did not reveal any such features, and the range over which recognition could be performed was smaller than the previous methods.

People are complex moving 3D self occluding shapes and therefore there is no simple holistic transformation between 2D silhouettes at different views. However, using knowledge that we are observing people walking, that some features, such as the leg motion, can be found and transformed between orientations (Chapter 6). To do this a novel model based gait recognition algorithm was developed, and shown to work, albeit with low recognition rates.

Several improvements were identified and implemented to the algorithm in the previous chapter (Chapter 7). The results of these improvements was a much smaller relative drop

in performance between orientations, however the algorithm has poor base recognition rate.

## 8.2 Future work

Although it has been shown the new algorithm works, the range of orientations it works over is not known, although it is expected to be much greater than any silhouette based algorithm is capable of. The relative drop in performance is approximately half that of the average image algorithm, suggesting the range of orientations would be double. However, a systematic error in the raw data was discovered that adversely effects inter orientation recognition. Additionally the recognition rate with artificial data does not decay linearly with the difference in orientation so the estimate of double could be too low or too high. The algorithm needs testing on range of real life data to understand its performance but at the time of writing no such database exists.

The within angle recognition rates, where data recorded at identical orientations is compared, are not as good as those achieved by other algorithms. There are several improvements that could be made to increase the accuracy of the model's alignment with the silhouette sequence.

The foot moves with respect to the lower leg, when the rear foot is in contact with the floor and before it is lifted for the forward swing, it is maximally raised. As the foot is brought forward it flexes down. The templates used do not take this into account and it is possible this is a source of error in our estimation of the lower leg angle.

The foot could be modelled as a cylinder, or as a template, however in some situations it is difficult to see what should be leg and what should be foot. A more robust solution may be to extend the templates themselves. In stead of using the mean silhouette, PCA could be performed upon the segmented legs and optimise along the major components with the other model parameters. Another approach would be to model the changing shape of the foot using an active shape model (Cootes et al., 1995).

Parallel projection are used to draw our silhouettes, some measures were taken to improve this, allowing each leg separate parameters and templates. This enables the further leg to appear smaller however this too could be a source of error in our measured angles. This could be solved by moving to a perspective projection.

Only camera views at different positions around the subject have been tested, to be fully orientation independent the algorithm would need to be tested and perhaps modified to make use of views encountered at different elevations. At other orientations the appearance of the foot may change more through a gait cycle, this would further necessitate modification of the foot templates.

As a subject passes the camera their distance between the camera and their relative orientation changes, templates would need to be modified as the subject moves.

All the work in this thesis has focussed on silhouettes, however there may be an accuracy gain by incorporating colour information from the original RGB images. By extracting colour templates of the lower leg more accurate alignments may be achievable, however there would be problems with shadows and moving clothing as well as the partial occlusion of one of the legs.

Using a real world based coordinate system with the floor modelled would provide additional cues when attempting to accurately segment out the feet

Consistent accurate motion estimates have been extracted, one might ask if these can be improved much more. The limit of the ability of shin motion to identify people may be being approached. An interesting set of experiments to conduct would be to align the templates with the real data as using the new template generation, then average the artificial silhouettes according to the average silhouette algorithm, for comparison with the results of the algorithm applied to the real data. A comparison with the recognition ability of the static templates and the recognition ability of the motion would be possible.

Background subtraction is always important for silhouette based algorithms, but can be extremely difficult in some types of scene. No experiment was performed to see how the model coped with different types of noise in the silhouettes. Model based algorithms such as the one described in this thesis have an advantage over silhouette based methods as it becomes possible to identify some sorts of errors in background subtraction. For example, the motion shape and existence of a subjects head is relatively consistent between frames.

Our proposed algorithm has successfully recognised subjects between distinct orientations, there are several directions research could take to improve these results.

# Appendix A

## Baseline Algorithm Results

### A.1 Within Orientation Confusion Matrices

The confusion matrices in Figure A.1 show the similarity between feature vectors recorded at the same orientation.

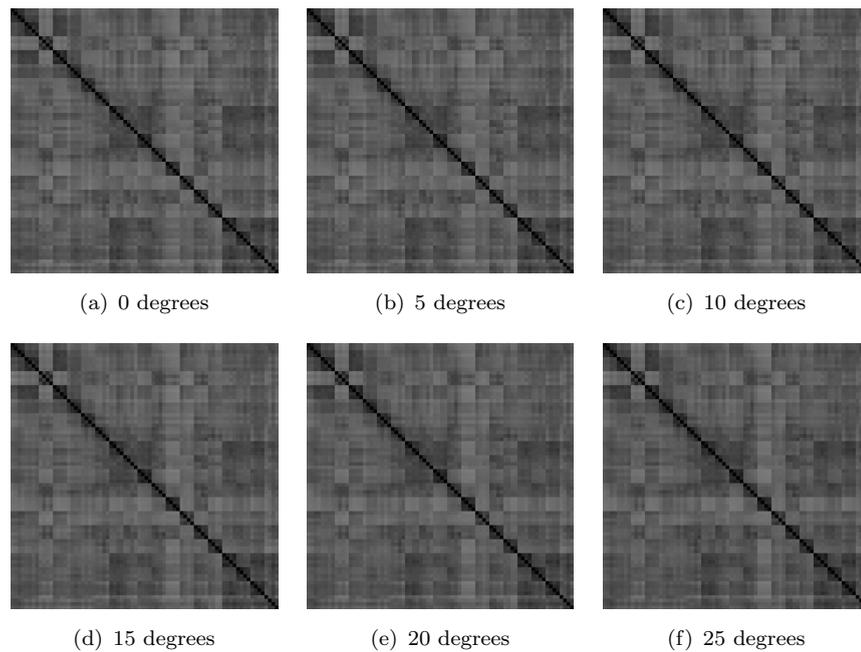


FIGURE A.1: The confusion matrix of all runs of all subjects at 0,5,...,90 degrees. 4 by 4 dark boxes along the diagonal indicate there is little difference between instances of the same subject, white areas show large difference between subjects.

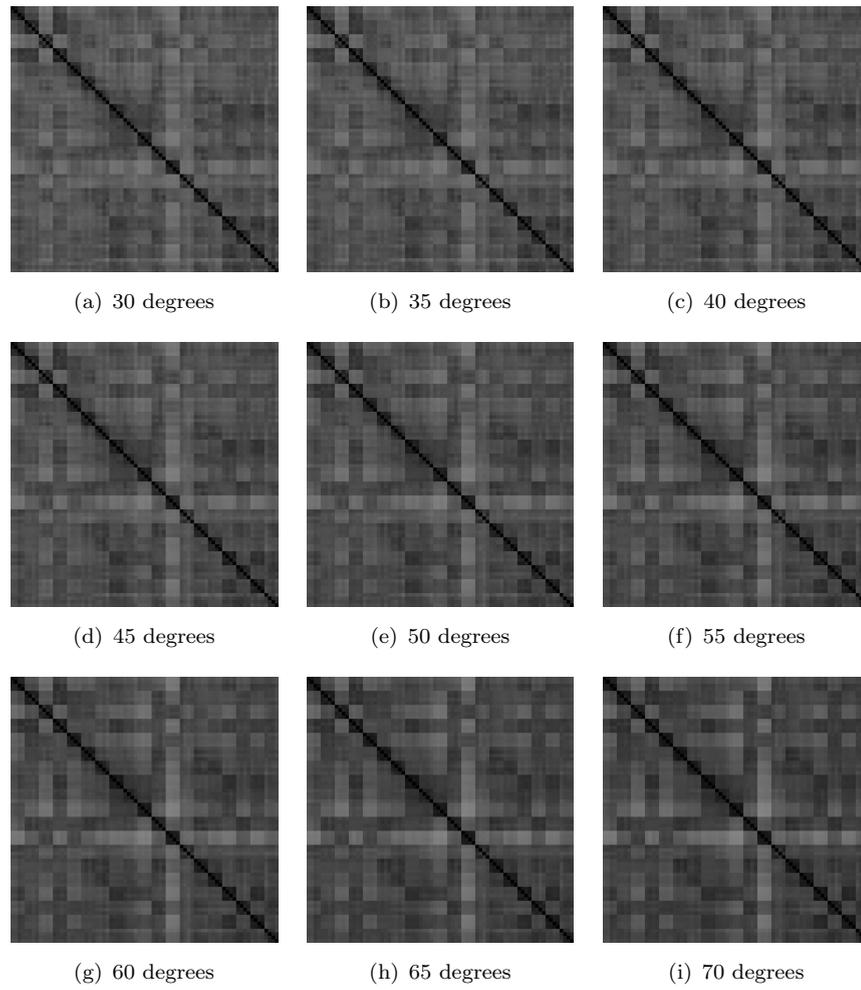


FIGURE A.2: Figure A.1 continued

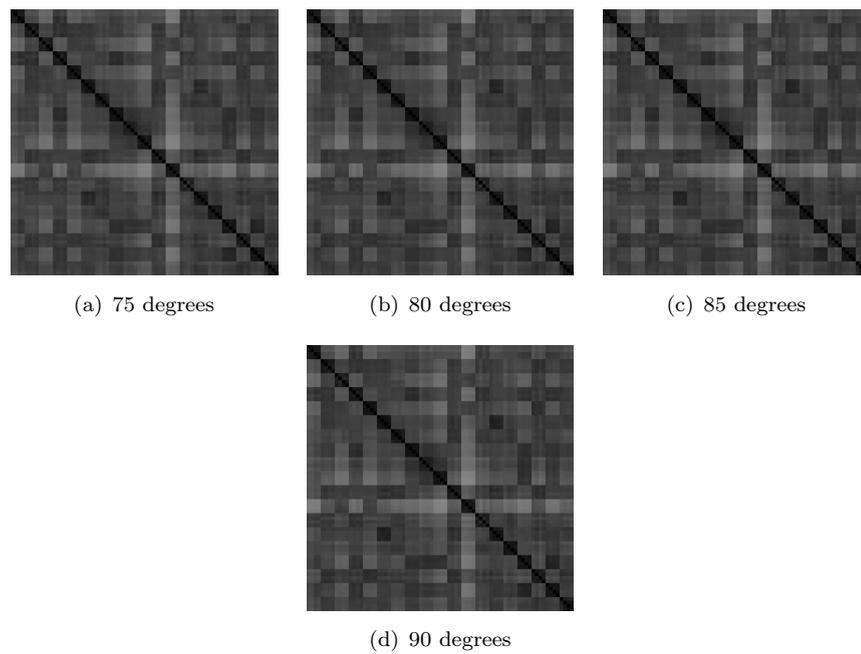


FIGURE A.3: Figure A.1 continued

## A.2 Between Orientation Confusion Matrices

The confusion matrices in Figure A.4 show the similarity between feature vectors recorded at 0 degrees and another orientation.

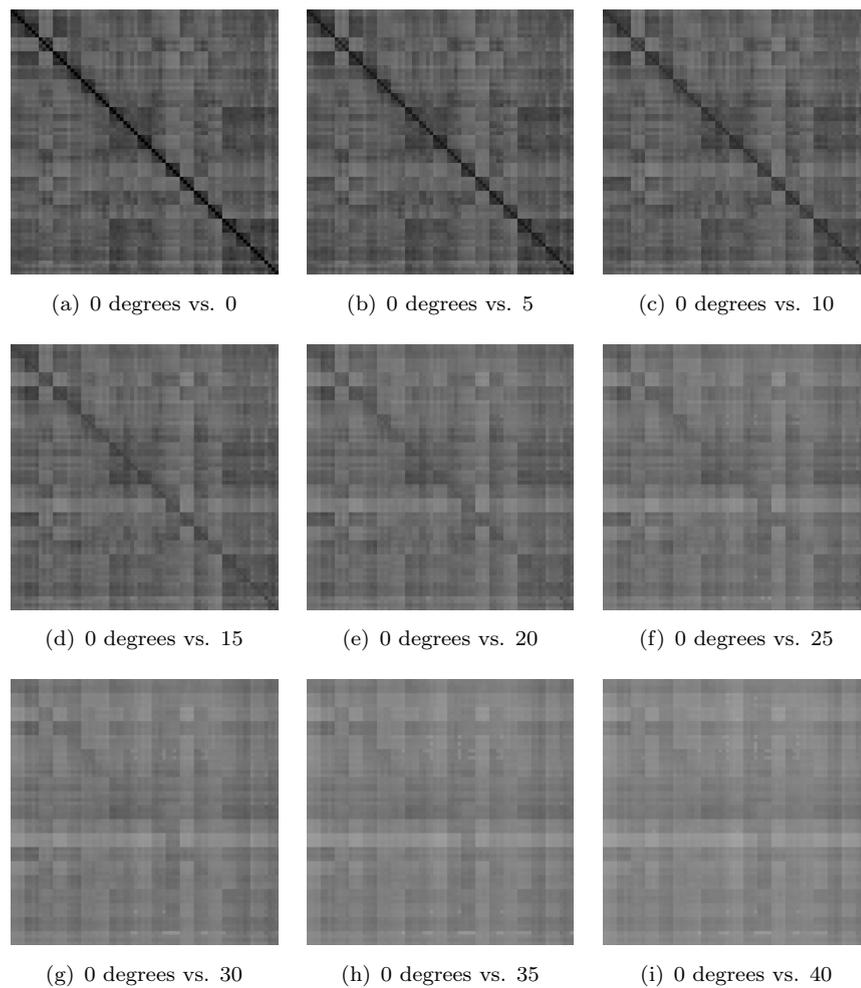


FIGURE A.4: Confusion matrices of feature vectors recorded at 0 degrees vs other orientations.

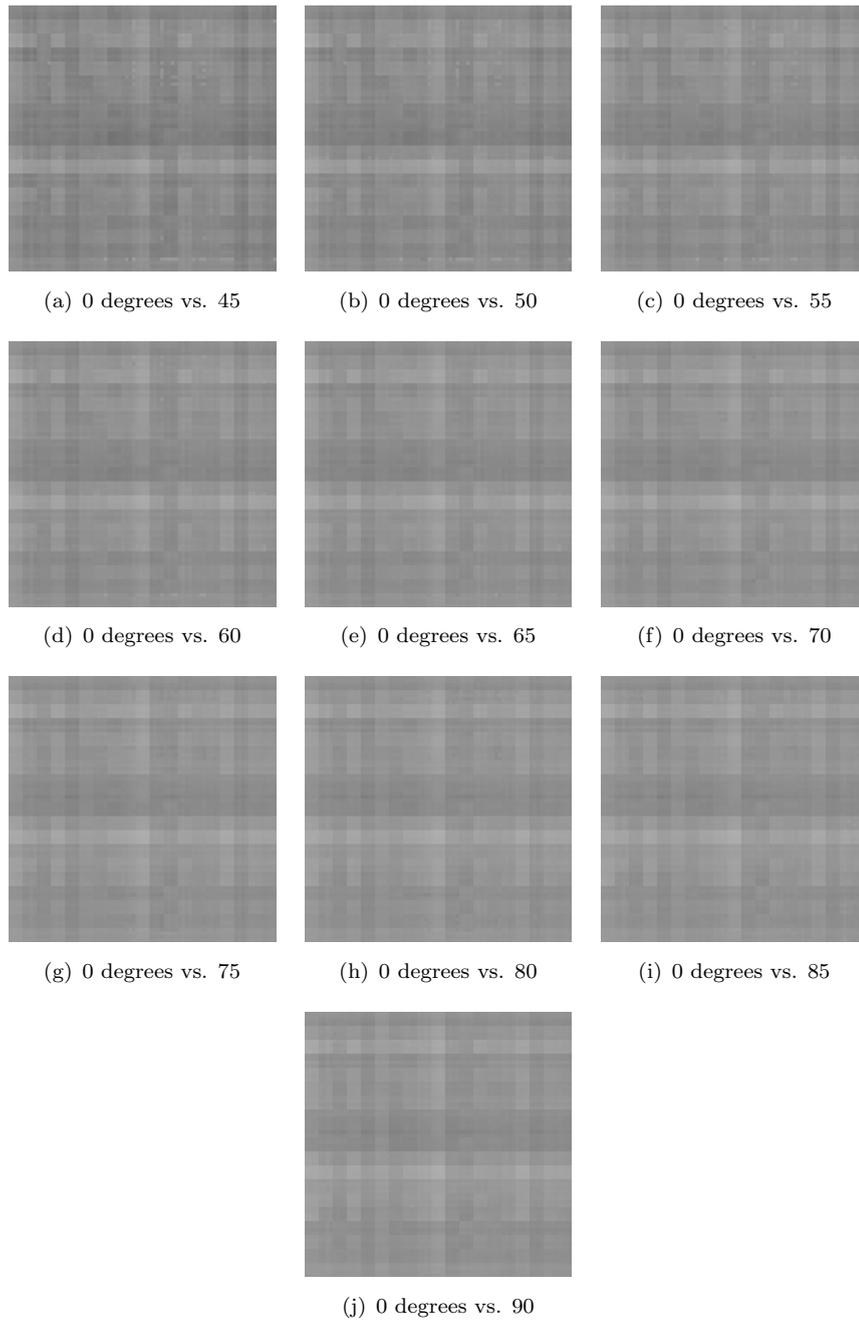


FIGURE A.5: Figure A.4 continued.

## Appendix B

# Average Silhouette Algorithm Results

### B.1 Within Orientation Confusion Matrices

The confusion matrices in Figure B.1 show the similarity between feature vectors recorded at the same orientation.

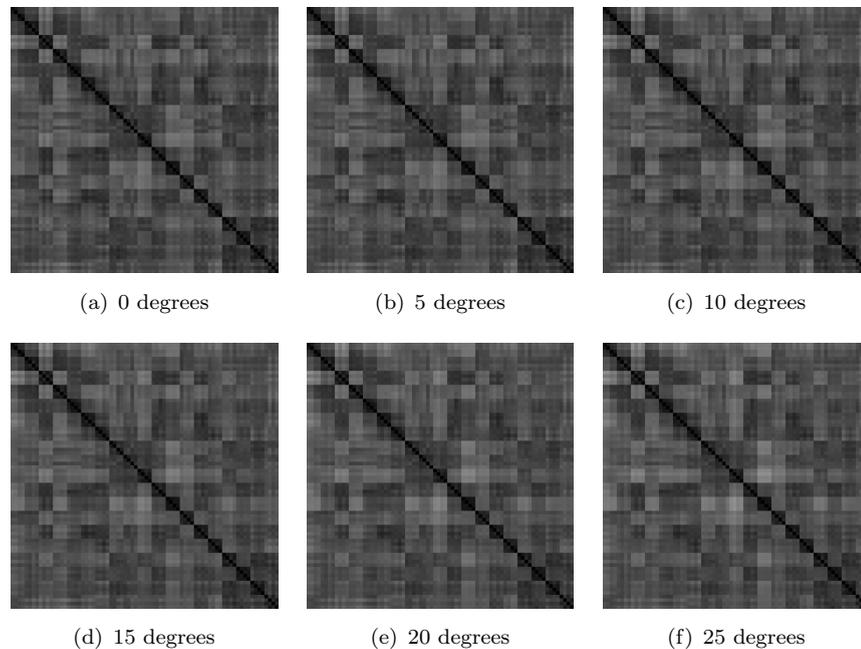


FIGURE B.1: The confusion matrix of all runs of all subjects at 0,5,...,90 degrees. 4 by 4 dark boxes along the diagonal indicate there is little difference between instances of the same subject, white areas show large difference between subjects.

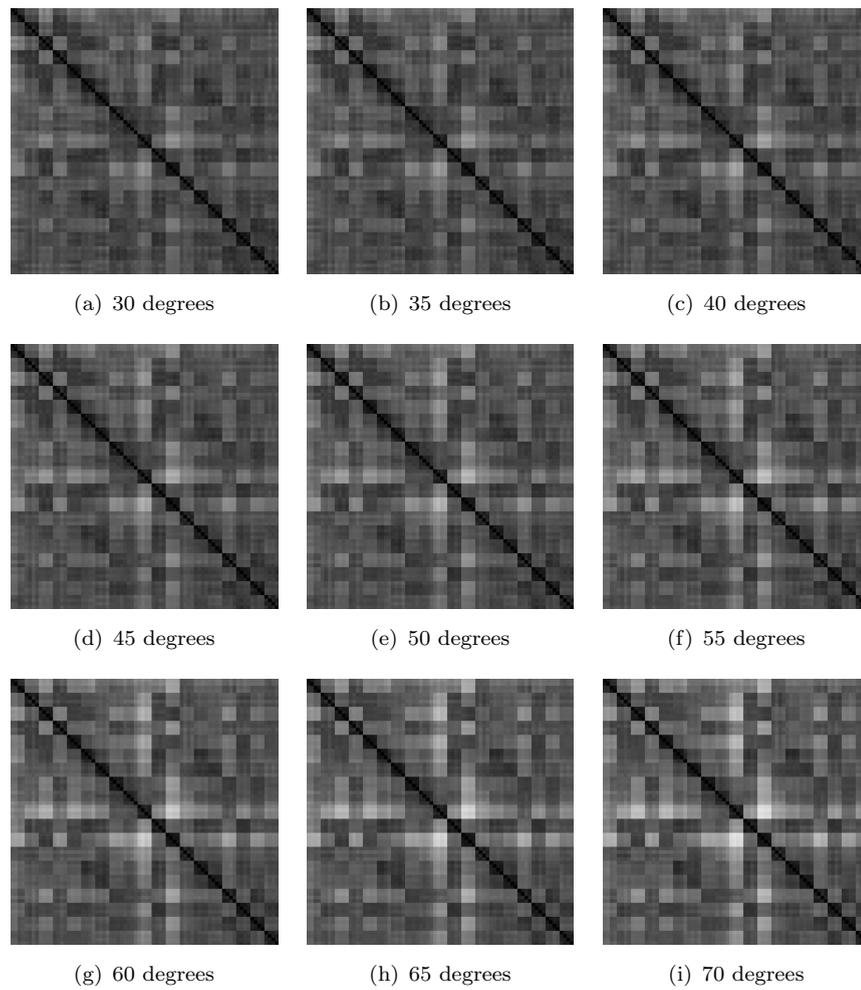


FIGURE B.2: Figure B.1 continued

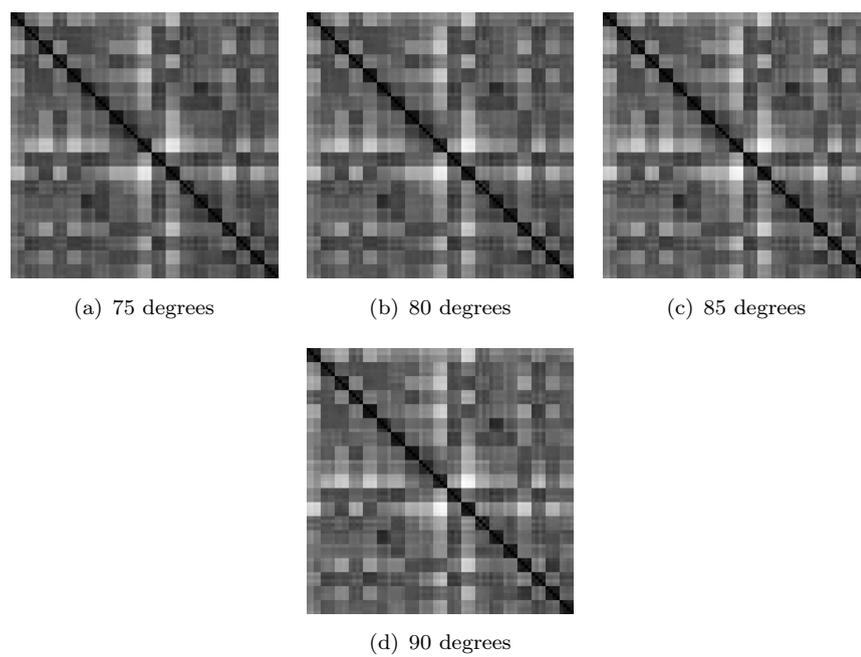


FIGURE B.3: Figure B.1 continued

## B.2 Between Orientation Confusion Matrices

The confusion matrices in Figure B.4 show the similarity between feature vectors recorded at 0 degrees and another orientation.

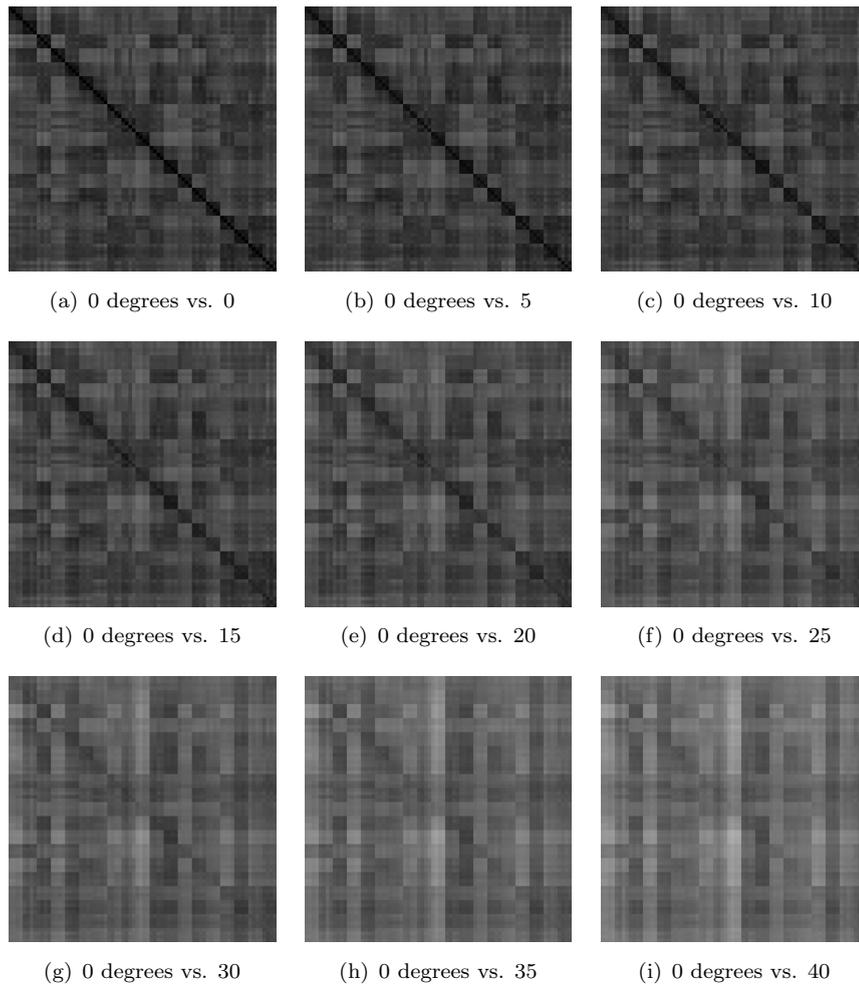


FIGURE B.4: Confusion matrices of feature vectors recorded at 0 degrees vs other orientations.

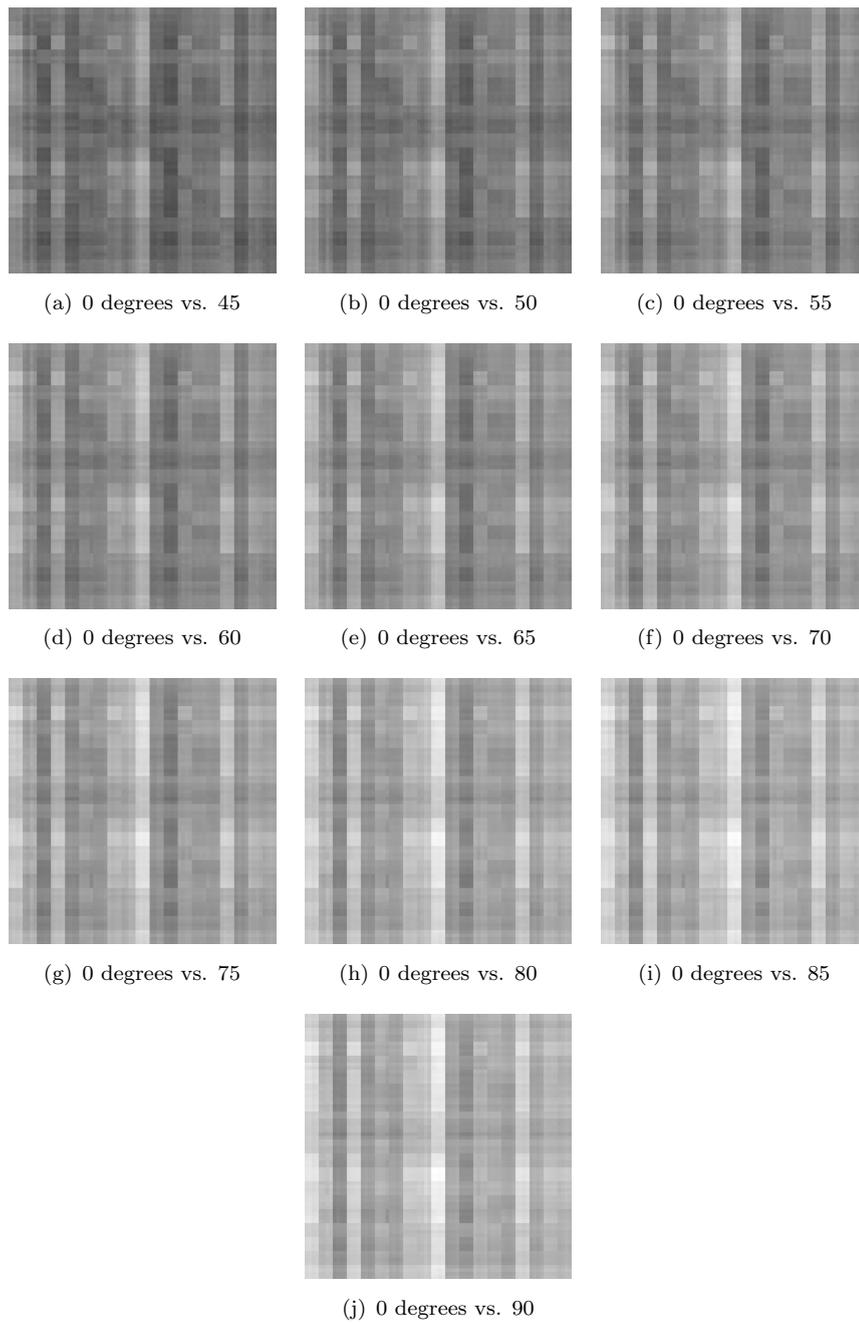


FIGURE B.5: Figure B.4 continued.

# Bibliography

- C. BenAbdelkader, R. Cutler, and L. Davis. Stride and cadence as a biometric in automatic person identification and verification. In *In Proceedings of IEEE Face and Gesture Recognition*, 2002a.
- Chiraz BenAbdelkader, Ross Cutler, and Larry Davis. Motion-based recognition of people in eigengait space. In *5th International Conference on Automatic Face and Gesture Recognition*, 2002b.
- J. Carter and M. Nixon. On measuring trajectory-invariant gait signatures. In *Proc. XIX Congress of the International Society for Photogrammetry and Remote Sensing*, pages 114–121, 2000a.
- John Carter. Average image, left vs right. Personal Correspondence, 2008.
- John N. Carter and Mark S. Nixon. On measuring trajectory-invariant gait signatures. In *Proc. XIX Congress of the International Society for Photogrammetry and Remote Sensing*, pages 114–121, 2000b.
- Upendra V. Chaudhari and Ganesh N. Ramaswamy. Speaker verification using target and background dependent linear transforms and multi-system fusion. In *In Proc. of EUROSPEECH-01*, 2001.
- Amit K. Roy Chowdhury, Amit Kale, and Rama Chellappa. Video synthesis of arbitrary views for approximately planar scenes. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2003.
- R Collins, R Gross, and J Shi. Silhouette-based human identification from body shape and gait. In *International Conference on Face and Gesture*, pages 351–356, October 2002.
- T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models—their training and application. *Comput. Vis. Image Underst.*, 61(1):38–59, 1995. ISSN 1077-3142.
- D. Cunado. *Automatic Gait Recognition via Model-Based Moving Feature Analysis*. PhD thesis, University of Southampton, 1999.

- David Cunado, Jason M. Nash, Mark S. Nixon, and John N. Carter. Gait extraction and description by evidence-gathering. In *Proceedings of the Second International Conference on Audio- and Video-Based Biometric Person Authentication*, pages 43–48, 1999.
- JE Cutting and LT Kozlowski. Recognizing friends by their walk: gait perception without familiarity cues. *Bulletin of the Psychonomic Society*, 9(5):353–356, 1977.
- Sarat C. Dass and Anil K. Jain. Fingerprint-based recognition, 2006.
- John Daugman. How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology*, 14:21–30, 2004.
- J.P. Foster, M.S. Nixon, and A. Prugel-Bennett. New area based measures for gait recognition. In *Proceedings Audio- and Video-Based Biometric Person Authentication*, pages 312–317, 2001.
- John Geisheimer. Gtri motion capture database. Personal Correspondence, 2001.
- Ralph Gross and Jianbo Shi. The cmu motion of body (mobo) database. Technical Report CMU-RI-TR-01-18, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, June 2001.
- James Hayfron-Acquah, Mark S. Nixon, and John N Carter. Human identification by spatio-temporal symmetry. In *Proceedings International Conference on Pattern Recognition*, volume 1, pages 632–635, Quebec, 2002.
- James B. Hayfron-Acquah, Mark. S. Nixon, and John N. Carter. Automatic gait recognition by symmetry analysis. In *Audio-and-Video-Based Biometric Person Authentication*, pages 272–277, 2001.
- G. Johansson. Visual perception of biological motion and a model for its analysis. In *Perception and Psychophysics*, pages 201–211, 1973.
- Johnson and Bobick. A multi-view method for gait recognition using static body parameters. In *3rd International Conference on Audio- and Video Based Biometric Person Authentication*, pages 301–311, June 2001.
- A. Kale, R. Chowdhury, and R. Chellappa. Towards a view invariant gait recognition algorithm. In *IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS03)*, 2003a.
- A Kale, N Cuntoot, B Yegnanarayana, A N Rajagopalan, and R Chellappa. Gait analysis for human identification. In *Audio, Video and Biometric Person Authentication*, pages 706–714, 2003b.

- Amit Kale, A.N. Rajagopalan, Naresh Cuntoor, and Volker Kruger. Gait based recognition of humans using continuous hmms. In *International Conference on Face and Gesture Recognition*, 2002.
- T. Lake. National dna database annual report for 2005-06, 2007.
- L. Lang and G. Leopold. Digital identification: Its now at our fingertips. *Eetimes*, 947, 1997.
- V. Laxmi, J. N. Carter, and R. I. Damper. Biologically-motivated human gait classifiers. In *Third IEEE Workshop on Automatic Identification Advanced Technologies*, pages 17–22, 2002.
- L. Lee. Gait dynamics for recognition and classification. In *MIT AIM*, 2001.
- James J. Little and Jeffrey Boyd. Recognizing people by their gait: The shape of motion. *Videre*, 1(2):1–32, 1998.
- Yanxi Liu, Robert Collins, and Yanghai Tsin. Gait sequence analysis using frieze patterns. Technical Report CMU-RI-TR-01-38, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, December 2001.
- Wojciech Matusik, Chris Buehler, Ramesh Raskar, Steven J. Gortler, and Leonard McMillan. **Image-based visual hulls**. In Kurt Akeley, editor, *Siggraph 2000, Computer Graphics Proceedings*, pages 369–374. ACM Press / ACM SIGGRAPH / Addison Wesley Longman, 2000.
- MP Murray. Gait as a total pattern of movement. *American Journal of Physical Medicine*, 46(1):290–333, Feb 1967.
- J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7:308–313, 1964.
- Mark S. Nixon, Tieniu N. Tan, and Rama Chellappa. *Human Identification based on Gait*. International Series on Biometrics. Springer, [”lib/utills:month\_11573” not defined] 2005.
- A. Niyogi, Sourabh and H. Adelson, Edward. **Analyzing and recognizing walking figures in XYT**. Technical Report 223, 1993.
- Mauricio Orozco-Alzate and César Germán Castellanos-Domínguez. I-TECH Education and Publishing, Vienna, Austria, June 2007.
- P. J. Phillips, S. Sarkar, I. Robledo, P. Grother, and K. W. Bowyer. The gait identification challenge problem: Data sets and baseline algorithm. In *ICPR*, pages 385–389, 2002a.

- P. Jonathon Phillips, Sudeep Sarkar, Isidro Robledo, Patrick Grother, and Kevin Bowyer. Baseline results for the challenge problem of human id using gait analysis. In *Proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition (FGR 2002)*, May 2002b.
- William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical recipes in C (2nd ed.): the art of scientific computing*. Cambridge University Press, New York, NY, USA, 1992. ISBN 0-521-43108-5.
- S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 3(52), 1965.
- J. D. Shutler, M. G. Grant, M. S. Nixon, and J. N. Carter. On a large sequence-based human gait database. In *Proc. 4th International Conference on Recent Advances in Soft Computing*, pages 66–71, December 2002.
- J. D. Shutler and M. S. Nixon. Zernike velocity moments for the description and recognition of moving shapes. In *Proc. British Machine Vision Conference*, pages 705–714, 2001.
- N. Spencer and J. N. Carter. Towards pose invariant gait reconstruction. In *IEEE International Conference On Image Processing*, pages 261–264, 2005.
- Nicholas M. Spencer and John N. Carter. Viewpoint invariance in automatic gait recognition. In *Proceedings of Third IEEE Workshop on Automatic Identification Advanced Technologies AutoID'02*, pages 1–6, 2002.
- Rawesak Tanawongsuwan and Aaron Bobick. Gait recognition from time-normalized joint-angle trajectories in the walking plane. In *IEEE Computer Vision and Pattern Recognition Conference*, December 2001.
- R. Veltkamp and M. Hagedoorn. **State-of-the-art in shape matching**. Technical Report UU-CS-1999-27, Utrecht University, the Netherlands, 1999.
- R. C. Veltkamp. **Shape matching: similarity measures and algorithms**. In *Proceedings Shape Modelling International*, pages 188–199. IEEE Press, 2001.
- G. Veres, L. Gordon, J. Carter, and M. Nixon. What image information is important in silhouette-based gait recognition. In *In Proceedings of IEEE Computer Vision and Pattern Recognition conference, Washington*, 2004.
- L. Wang, T. Tan, H. Ning, and W. Hu. Fusion of static and dynamic body biometrics for gait recognition. In *IEEE Transactions on Circuits and Systems for Video Technology Special Issue on Image- and Video-Based Biometrics*, 2004.

---

ChewYean Yam, Mark S. Nixon, and John N. Carter. Extended model-based automatic gait recognition of walking and running. In *Proceedings of 3rd International Conference on Audio and Video Based Biometric Person Authentication, AVBPA 2001*, pages 278–283, 2001.