UNIVERSITY OF SOUTHAMPTON

Automatic Gait Recognition via Area Based Metrics

Jeffrey Paul Foster

Doctor of Philosophy

Faculty of Engineering and Applied Science

Department of Electronics and Computer Science

Submitted March 2003

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING AND APPLIED SCIENCE

DEPARTMENT OF ELECTRONICS AND COMPUTER SCIENCE

Doctor of Philosophy

AUTOMATIC GAIT RECOGNITION VIA AREA BASED METRICS

by Jeffrey Foster

Gait is a new biometric aimed at recognising a subject by the way they walk. Gait differs from traditional biometrics in that it is a function of both space and time. We present two new approaches for automatic gait recognition.

The first of these two approaches, area masks, aims to recognise a subject by their dynamics of area change within specific regions of the image. This approach focuses on the temporal nature of gait, which has been neglected by previous statistical approaches. The second approach, moment based descriptors, describes a shape in terms of geometric invariants. A family of shape descriptors is formed by using a masking circle proportional to the area of the image. We show that a family of descriptors is a better discriminator than using just simple moments alone.

We use a state of the art database, which is currently the largest available, and present extensive experimental results examining gait as a biometric, gender discrimination, gait symmetry and performance analysis. A simple nearest neighbour classifier is used to discriminate between subjects and this provides a measure of baseline performance.

Our new approaches provide promising results on the largest available database. Future work will concentrate on extending the approaches to deal with gait filmed under real-world conditions.

Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my supervisors Mark Nixon and Adam Prügel-Bennett for their support and input during my research. I would also like to thank my wife, Charlotte, for reading through my work hundreds of times and introducing eigenvectors as an expletive.

I also gratefully acknowledge partial support by the European Research Office of the US Army under Contract No. N68171-01-C-9002, and support from the EPSRC.

Chapter 1 Context and Contributions

1.1 Introduction

Biometrics are a featured characteristic of people that allows them to be distinguished from others. Examples include face, fingerprints, hand geometry or irises. In today's increasingly security conscious society, biometrics are of growing importance for authentication, authorisation and security.

Essentially, there are two phases in any biometric system. The enrolment phase consists of capturing the biometric (via video camera or similar sensors) and turning this information into a set of training data by finding a characteristic of the captured data that distinguishes between subjects. The second stage is comparing a new captured sample to the set of subjects in the training database using a classification algorithm.

In recent years, with advances in computing power, gait has become a potential practical biometric. Gait is a biometric aimed at recognising subjects by the manner in which they walk. Gait is a new challenge in the field of biometrics, because it is both a function of space and time. Gait has notable advantages over other biometrics in that it requires no subject contact, is potentially better discernable from large distances than other biometrics and is less likely to be obscured.

1.2 Scope of Thesis

This thesis presents a contribution to automatic gait recognition. For the purposes of this thesis, we shall assume gait to be filmed under the following conditions:

- A single static camera and only one walking subject appears in any scene

By limiting ourselves to a single subject in the view at any one time, the task of silhouette extraction is greatly simplified. A single static camera is used since we are concerned solely with the movement of the human body and this is significantly easier to isolate if a single static camera is used.

- Each subject is walking normal to the camera's plane of view.

a subject is walking normal to the camera's plane of view, the motion of the legs is clearest. This avoids scaling problems during silhouette extraction if the subject was walking towards or away from the camera. From the video clip of a subject walking, we extract windowed binary silhouettes. Binary silhouettes are used because they dramatically reduce the dimensionality of the data, and colour information contains no gait information.

To discriminate between subjects, a feature vector needs to be generated for each subject that is similar for different sequences of the same subject, but dissimilar for different subjects. This feature vector can then be used to discriminate between the subjects, using a classification algorithm.

In this thesis, we present two new contributions to automatic gait recognition, together with extensive experimental results. Our first new technique, area masks, aims to describe a gait sequence by the dynamics of area change within specific regions of the silhouette. Our second technique, moment based descriptors, describes a gait sequence in terms of geometric invariants.

The baseline property that is being described by both techniques is area. Our first technique, area masks, measures raw area in certain portions of the image. By utilising the temporal component of gait, we show that there is a large amount of information contained in area data. The primary advantage of this technique is one of speed. As a baseline metric, merely measuring area, the computational complexity of the approach is extremely low allowing us to undertake many experiments. The second technique, moment based descriptors, describes a gait sequence not just in terms of the area but also in terms of other geometric properties. We use Hu [50] invariant moments to describe a sequence and are thus invariant to rotation, translation and scaling.

Extensive experimental results are presented which use the SOTON database [1]. The SOTON database consists of 114 subjects with a minimum of eight samples each, filmed under laboratory conditions and, at the time of writing, is the largest gait database available.

Publications related to this thesis are listed before the References.

1.3 Thesis Overview

The remainder of this thesis is arranged as follows.

- Chapter 2: Gait Recognition

Chapter 2 reviews previous work relevant to automatic gait recognition. Medical and psychological research is presented which supports the notion of gait as a biometric. Previous work on automatic gait recognition is presented with specific emphasis on statistical approaches. In conclusion, this chapter summarises what has been learnt from previous work and outlines the aims of this research.

- Chapter 3: Gait Recognition via Area Based Metrics

Chapter 3 introduces area masks for automatic gait recognition. Essentially, area masks isolate portions of the image and measure the area change within these regions. They are invariant to vertical symmetry and we show that it is possible to compensate for changes in the angle of the walker relative to the camera. Area masks provide both temporal and static information.

- Chapter 4: Gait Recognition via Moment Based Descriptors

Chapter 4 demonstrates how moment based descriptors can be used for automatic gait recognition. Moment based descriptors provide a geometric description of the shape that is invariant to rotation, translation and scaling. A family of shape descriptors is formed by using masking circles proportional to the size of the image. By using a family of shape descriptors, the quality of description improves and thus better discrimination is possible between subjects.

- Chapter 5: Automatic Gait Recognition

Chapter 5 presents experimental results when using gait as a biometric. The new SOTON database is used and results are presented from both area masks and moment based techniques. We also examine in detail where the recognition performance is derived. Are we recognising subjects by variation in their dynamic cues (i.e. manner of walking) or by differences in their body shape?

- Chapter 6: Gender Discrimination

Medical studies have shown that there is a difference between the gait of male and female subjects, with males tending to swing their shoulders from side to side more than their hips, and females tending to swing their hips more than their shoulders. In this chapter, we examine whether moment based descriptors and area masks are able to fulfil the task of gender discrimination.

- Chapter 7: Is Gait Symmetric?

Gait has long been considered a symmetrical pattern of moment, and we have assumed this when using the SOTON database. Recent research has suggested that gait may have an asymmetric component and in this chapter, we examine this by considering the heel strikes in a gait cycle and the direction of travel.

- Chapter 8: Performance Analysis

This section presents performance analysis of both area masks and moment based descriptors. Noise is added to the database and the affects of this on recognition performance are examined. With the area-based technique, the subject is required to be centred in a windowed image so we examine the affects of the subject being misaligned in the window. We examine the invariance properties of the moment-based technique by using rescaled and rotated silhouettes. Finally, we examine the effects of using low-resolution silhouettes with both techniques.

- Chapter 9: Intra-Subject Variance

Preliminary results are presented from the new intra-subject variance database which aims to address how much an individual subjects gait differs over a variety of circumstances. The chapter illustrates that by using the temporal components of gait we can gain advantages when the subject's silhouette is changed (i.e. by changing clothing or by load carrying).

- Chapter 10: Conclusions and Future Work

This chapter summarises the work presented in this thesis and discusses potential problems of area masks and moment based descriptors. Future work aimed at extending and improving these techniques for automatic gait recognition is presented.

Chapter 2 Gait Recognition

2.1 Introduction

It is not immediately apparent that gait can be used as a biometric. However, even from Shakespearian times there has been reference to gait recognition. In 'The Tempest' [Act 4, Scene 1], Ceres observes "High'st Queen of state, Great Juno comes; I know her by her gait". The Oxford English dictionary describes gait as "a manner of walking or forward motion".

Gait has several notable advantages over other biometrics such as fingerprints. It is a non-invasive technique meaning that the subject need not know they are being recognised. Gait also allows recognition from a great distance where other biometrics, such as face recognition, may fail. Bank robbers can disguise their gait less easily than their face; in fact, disguising one's gait only has the effect of making oneself look more suspicious!

Medical studies support the notion of gait as a biometric. Indeed, gait is more than that and can be used to diagnose a wide variety of medical conditions [2-4] from neuro-muscular disorders to congenital joint defects. Psychological studies again support the hypothesis that gait can be used for recognition. Studies using moving light displays have indicated that subjects can be recognised solely by the manner in which they move, rather than recognition by silhouette shape. A study by Stevenage et al. [5] showed that humans can recognise gait even under adverse conditions. In recent years, with advances in computing hardware, gait has become a potential practical biometric and techniques have been developed and tested on small databases.

This chapter reviews previous work relevant to this thesis. Section 2.2 reviews medical studies by Murray [6, 7] which concern the uniqueness of gait. Section 2.3 looks at psychological studies by Johansson [8], Cutting and Kozlowski [9] and Stevenage [5]. Section 2.4 looks at recent work on automatic gait recognition, with particular emphasis placed on statistical approaches. Section 2.5 looks at model-based approaches to automatic gait recognition. Section 2.6 examines statistical based approaches to automatic gait recognition. In conclusion, section 2.7

discusses what can be learnt from previous approaches to automatic gait recognition and discusses the new work this thesis will present.

2.2 Medical Studies of Gait

Murray's work [6] considered gait as a total pattern of movement. Murray analysed the walking patterns of sixty normal men aged from 20 to 65 years old. Twenty simultaneous gait components were measured including serial displacement patterns of the head, neck, trunk and upper and lower limbs. Murray described gait throughout a total walking cycle. A total walking cycle is defined as the time interval between successive heel strikes of the same foot. The measurements were filmed by means of interrupted-light photography. The subjects, with reflective targets secured to specific anatomical landmarks, walked before a camera in the illumination of a strobe-light flashing at 20Hz. A mirror was used to note aspects of the overhead view and this was recorded onto film.

Murray et al [7] looked at the walking patterns of normal men. Sixty subjects in five age groups (20-25, 30-35, 40-45, 50-55 and 60-65) were selected and each subject considered themselves normal and had normal strength and range of motion. The subjects were filmed as before and the individual gait components were recorded. Murray concluded that each movement pattern (in terms of measured components) was "strikingly similar for repeated trials of the same subject". Murray's work also suggests that if all gait movements are considered then gait is unique. This makes gait an ideal candidate for use as a biometric.

2.3 Psychological Studies of Gait

Psychological research suggests that humans have a remarkable ability to recognise and distinguish between different types of motion. Johansson [8] performed an experiment with moving light displays (MLD's). He attached small light sources at the major joints of human actors, and filmed them moving about in darkness. When the film was played to observers, they immediately interpreted the moving pattern of isolated points as a human figure, even when presented with only a few of these moving dots. This raised the question whether recognition of moving parts could be achieved directly from motion without structure recovery.

Psychological studies from Cutting and Kozlowski [9] showed how viewers can recognise themselves and others in an abstract display of their movements, independent of other familiarity cues such as shape and size. Highly reflective tape was wrapped around walking joints and subjects were filmed walking normal to the camera's plane of view. When asked how they recognised subjects, viewers tended to mention as clues certain critical features of the display, such as the speed, bounciness, rhythm of the walker, amount of arm swing or the length of steps.

In another study by Kozlowski and Cutting [10], they examined recognising the sex of a walker from a dynamic point light display. Subjects were filmed using a sagittal (side) view, which meant direct information about the width ratio of subjects was obscured. Gender discrimination performance was far from perfect, at 63% correct on average, which is not significantly better than chance (50%). In a later report, Mather and Murdoch [11] looked at the area of gender discrimination from biological motion displays. They showed that frontal or oblique views are much more effective than a side view for distinguishing between the genders of subjects, and emphasised the importance of hip and shoulder movements for gender judgements. By presenting subjects with different views with which to distinguish between gender, results improved to an accuracy of 79%.

Stevenage et al [5], in a more recent study, again confirmed the possibility of recognising people by their gait, but now using video. The study concentrated on determining whether illumination or length of exposure could impair the ability of gait perception. The study confirmed that, even under adverse conditions, gait could still be perceived. Psychological studies therefore clearly support the case for gait as a biometric.

2.4 Automatic Gait Recognition

Recently, there has been much attention devoted to using gait as a biometric [12-36]. With advances in computing power and cheaper mass storage devices, it is now possible to process large gait databases. As an example of the size of a typical gait database, the new SOTON database consists of 114 subjects and almost 60,000 individual images and is approximately 1 terabyte in size.

Approaches to automatic gait recognition can broadly be divided into two. Model based approaches [21, 36, 37] aim to explicitly model gait as a series of equations.

One of the main advantages of the model-based approaches is their handling of occlusion, which is of especial importance in gait as the human body is self-occluding when walking. However, the model used to describe gait is often complicated, resulting in a high computational cost.

Holistic / statistical approaches [12-14, 16, 17, 19, 20, 26, 31, 34, 38-40] aim to process a gait sequence to find a set of measurements to distinguish between subjects. The disadvantage of traditional statistical approaches is that they are not intimately related to gait and just produce raw numbers to distinguish between subjects. In addition, most current statistical techniques fail to take into account the temporal component of gait and instead focus on distinguishing between collections of silhouettes, rather than a sequence as a whole.

2.5 Model-Based Approaches to Gait Recognition

Cunado et al. [21, 22] developed a model based approach to automatic gait recognition. The gait signature is derived from the spectra of measurements of the thigh's orientation. The legs are considered as interlinked pendulums and the phase-weighted Fourier magnitude spectrum is used as the feature to distinguish between different subjects. The model proposed is simple, but provides enough dynamic information for recognition. The velocity Hough transform [41] was used to extract the motion of the legs via evidence gathering. The approach achieved a recognition rate of 90% on a database of ten subjects with four sequences each, using the *k*-nearest neighbour classifier.

Yam et al[35, 36] presents an analytical model of gait. Gait is modelled as a dynamic coupled oscillator that can be used to describe both walking and running motions. Leg motion is extracted using temporal template matching with a model defined by forced coupled oscillators. The biometric signature is then derived from Fourier analysis of the variation in the motion of the thigh and lower leg. Classification of subjects is performed using the *k*-nearest-neighbour classifier. Recognition rates are similar to that achieved by other techniques with a similarly sized database (over 90% on a database of twenty subjects).

Bhanu and Ham [37] adopts a model based on the geometric representation of each part of the human body. The approach estimates 3D human walking parameters by performing a least squares fit of the 3D kinematic model to the 2D silhouette extracted from using a single camera. The approach has the advantage that it is not reliant on the subject walking normal to the camera's plane of view. Using this approach, a recognition rate of 77% (out of 30 sequences tested) was achieved.

2.6 Holistic Approaches to Automatic Gait Recognition

2.6.1 Holistic Gait Descriptors

Little and Boyd's [14] approach aims to describe the motion of a moving human figure in order to recognise individuals by variation in the characteristics of the motion description. The algorithm starts with a short sequence of images of a moving figure (taken by a static camera) and from this dense optical flow data is derived. A range of scale-independent scalar features of each flow image is calculated that characterises the spatial distribution of the flow. These scalar features represent the shape of the motion. The scales are based on various moments of the set of moving points. To characterise the shape of the motion, rather than the shape of moving points, the points are weighted by the optical flow components.

Next, the periodic structure of these sequences of scales is analysed. All of these sequences share the same fundamental period of gait, but they differ in phase. Although there are several regularities in the relative phases of the signals, some phases show significant statistical variation. Therefore, vectors of phase measurements derived from each image sequence can be used to recognise individuals by the shape of their motion. The representation is model-free, and therefore could be used to characterise the motion of other non-rigid bodies.

2.6.2 Gait Recognition using Symmetry

Hayfron-Acquah et al. [25, 26, 42] have developed a method for automatic gait recognition based on analysing the symmetry of human motion, using the generalised symmetry operator. This operator locates features by their symmetrical properties rather than relying on the borders of a shape or general appearance.

As with previous approaches to gait recognition, the system uses binary silhouette templates derived from a video sequence of a subject walking. The symmetry operator uses an edge map of images in the sequence to assign symmetry magnitude and orientation to image points, accumulated at the midpoint of each pair of points. A symmetry feature to describe the entire sequence is then calculated as the average of each individual symmetry map.

Following calculation of the symmetry feature, the Fourier transform is applied to each signature and the transform is low pass filtered to reduce sensitivity to high frequency components. The k nearest neighbour algorithm was used for classification and the technique achieved high recognition rates (over 90%) on various databases. An advantage of this approach is that, because of the averaging of symmetry maps, the technique is very tolerant to missing frames.

2.6.3 Describing Motion using Zernike Velocity Moments

Shutler et al [32-34] proposed a generic approach to describe moving shapes. The approach aims to describe a shape not only by its shape but by its motion throughout the sequence. The technique uses an extended form of the established Zernike moment set [32] which is well proven in pattern recognition.

In application to gait recognition, the technique applies Zernike velocity moments to produce a compact invariant description of both shape and motion. Zernike velocity moments are calculated on a single gait cycle of silhouette templates. Features for classification were selected using an ANOVA technique. Classification used a k nearest neighbour algorithm and recognition rates were impressive (100% on a database of six subjects with seven sequences each). One of the main advantages of this approach is that the handling of occlusion is impressive.

2.6.4 Baseline Recognition from Body Shape and Gait

Collins et al [20] approach aims to establish a simple baseline method for human identification based on body shape and gait. Silhouette extraction is achieved using simple background subtraction and thresholding, followed by using a 3x3 median filter to remove isolated pixels. A direct comparison of body silhouette images is not possible since 2D body shape changes non-rigidly throughout the gait cycle as the limbs move. Each sequence of silhouettes is processed to extract key frames representing landmark poses within the gait cycle (such as heel strikes). These key frames are extracted using simple measures such as silhouette width.

After locating key frames in an image sequence, a template is constructed for each by scaling and cropping the key frame silhouette. Templates are labelled according to which key frame they belong. A training database is formed consisting of all key frame templates extracted from each training gait sequence. Templates from a test subject can be compared by means of a template-matching algorithm and a nearest neighbour algorithm is used for recognition.

The algorithm was tested on various databases, with the number of subjects ranging from 24 to 55. Recognition results of up to 100% were achieved. The approach implicitly captures biometric shape cues such as body height, width and body-part proportions, as well as gait cues such as stride length and amount of arm swing. The approach is limited to classifying test sequences taken from the same viewing angle as the training sequences. The approach does not implicitly take into account the temporal nature of gait and relies instead on the detection of key poses that are compared using template matching.

2.6.5 Baseline Recognition Results using Gait Analysis

Philips et al. [31] looked at the issues involved in gait recognition and developed a simple algorithm to recognise subjects. The baseline algorithm, which was designed to be simple and fast, is composed of three parts. The first part semi-automatically defines bounding boxes around the moving person in each frame of the sequence. The position of the start, middle and end boxes is inputted by the user, and the algorithm determines the placement of the remaining boxes.

The second stage of the algorithm extracts the subject from the background. Essentially this is accomplished using a simple background subtraction algorithm followed by filtering to improve the image quality.

The final part of the algorithm computes the correlation between any two sets of silhouettes and allows classification to take place. Frame similarity for two frames is simply defined as the ratio of the number of pixels in their intersection (i.e. common to both frames) to the number of pixels in their union (a measurement of area). The correlation is simply defined as the sum of frame similarity measures throughout an entire sequence. Recognition results using this baseline algorithm were surprisingly good achieving a recognition rate of over 70%.

2.6.6 Gait Analysis using Ellipsoids

Lee and Grimson [30] demonstrated how to represent gait using a simple ellipsoid model. As with other approaches, gait is considered to be a collection of binary silhouettes. The silhouette is divided into seven regions. For each of these seven regions, a silhouette is fitted. The features extracted from each of these regions are the centroid, aspect ratio of major and minor axis of the ellipse and the orientation of the major axis of the ellipse.

Using these features, classification can take place. The algorithm was tested on a database of 25 subjects and recognition rates of up to 100% were achieved. The approach is view dependent and if clothing changes dramatically then the quality of ellipse fitting will be greatly reduced.

2.6.7 Motion-based Recognition of People in Eigenspace

BenAbdelkader et al. [18] proposed an approach based on self-similarity plots. The approach maps a sequence of images of a walking person to a similarity plot, which is defined as the matrix of self-similarities between each pair of images in the sequence.

As in previous approaches, the first stage is segmenting the silhouette from the background and storing this information as simple binary silhouette templates. Following this a sequence consisting of *N* frames is analysed for each subject and the similarity plot is computed. Details on how this is calculated can be found in [18, 19]

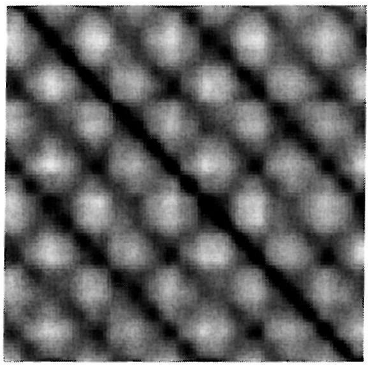


Figure 1 Sample Self -Similarity Plot for Gait Sequence

The self-similarity plot has useful properties related to gait. The intersections of the off-diagonals and cross diagonals encode the frequency and phase information about gait. Specifically, each intersection corresponds to a key pose in the gait cycle (e.g. legs furthest apart or legs closest together).

Following the calculation of the similarity plot, principal component analysis (PCA) is used to reduce the dimensionality of the feature and thus make classification easier. Following PCA, we are left with an "eigengait" which is the signature used to classify subjects. Using standard pattern recognition algorithms (k nearest neighbour) a recognition rate of 72% was achieved using a database of 25 people consisting of fifty sequences.

2.6.8 Automatic Gait Recognition via PCA + CA

The approach of Huang et al[27, 28] aims to recognise gait by using a combination of principal component analysis (PCA) and canonical analysis (CA). PCA is used to reduce the dimensionality of the data and then CA (also known as linear discriminant analysis) is used to separate the classes and perform classification.

Principal Component Analysis reduces the dimensionality of the data by finding the directions of maximum variation within the dataset and using a subset of these directions (eigenvectors) to project the data into a low dimensional subspace that represents the image with minimum error.

Huang's approach used binary silhouettes of a walking subject. We shall survey Huang's approach in more details than others as some of its basic statistical notions will find use in our later research. These binary silhouettes were represented in the form of a vector. We assume that we have a set of classes C (representing a subject) and for each instance of the class $c \in C$ we have a sequence x_c represented as a collection of vectors $x_c(i)$, where i labels the time at which the image was taken. Each sequence x_c consists of N_c individual images and the total number of images in the entire training set is N.

The mean template image for the full training set is given by

$$m = \frac{1}{N} \sum_{c \in C} \sum_{i \in N_c} x_c(i) \tag{2-1}$$

By subtracting the mean from each template image, the entire training set can be described by a $n \times N$ matrix $\mathbf{X}$ with each image forming one column of $\mathbf{X}$ and n is the n is the number of pixels in an image.

$$\mathbf{X} = [x_1(1) - m, \ldots, x_1(N_1) - m, \ldots, x_c(N_c) - m] \tag{2-2}$$

Principal component analysis uses the eigenvalues and eigenvectors generated by the data covariance matrix to rotate the original data coordinates along the direction of maximum variance. Let the rank of the matrix $\mathbf{XX}^\mathrm{T}$ be K, then the K nonzero eigenvalues of $\mathbf{XX}^\mathrm{T}, \lambda_1, \ldots, \lambda_K$, satisfy the fundamental eigenvalue relationship where $i = 1, \ldots, K$ and $\mathbf{R} = \mathbf{XX}^\mathrm{T}$

$$\lambda_i e_i = \mathbf{R} e_i \tag{2-3}$$

In order to solve the equation above we need to calculate the eigenvalues and eigenvectors *(e)* of the $n \times n$ matrix $\mathbf{R}$. This is computationally intractable for typical image sizes (e.g. 64 x 64 images give n as 4096). Based on singular value decomposition [43] we can compute another matrix instead, $\mathbf{Q}$, which is of size $N \times N$ and much smaller than $n \times n$ in practical problems.

$$\mathbf{Q} = \mathbf{X}^\mathrm{T}\mathbf{X} \tag{2-4}$$

The eigenvalues ($\lambda'_1, \ldots, \lambda'_K$) and associated eigenvectors ($e'_1, \ldots, e'_K$)of the matrix $\mathbf{Q}$ are related to those in $\mathbf{R}$ by the following equations.

$$\begin{aligned} \lambda_i &= \lambda'_i \\ e_i &= \lambda_i^{-\frac{1}{2}} \mathbf{X} e'_i \end{aligned} \tag{2-5}$$

The eigenvectors are used as an orthogonal basis to span a new vector space. Using this basis, each template image can be projected to a single point in K-dimensional space. According to the theory of PCA, each image can be approximated by taking the $k \le K$ largest eigenvalues. This partial set of k eigenvectors spans an eigenspace in which $y_c(i)$ are the points that is a projection of the original image $x_c(i)$

$$y_c(i) = [e_1, \ldots, e_k]^T x_c(i) \tag{2-6}$$

After PCA has been performed, we are left with a low dimensional vector $y_c(i)$ that represents the original images. Each sequence that is projected into eigenspace forms a trajectory that represents a low dimensional representation of the original data.

Canonical analysis (also known as linear discriminant analysis) [44] is a procedure for finding a low-dimensional subspace which gives a high discrimination between classes of data. Canonical analysis is used to maximise the difference between classes whilst simultaneously minimising the distance within classes. It differs from principal component analysis in that it uses knowledge of the correct classification of the data points. It is therefore a supervised as opposed to an unsupervised learning algorithm. We assume that we have a set of classes C (representing a subject) and for each class $c \in C$ we have a set of instances y_c consisting of vectors $y_c(i)$, where i labels the instances (individual images). The class means are given by

$$\mu_c = \frac{1}{|y_c|} \sum_{i \in y_c} y_c(i) \tag{2-7}$$

Whilst the mean of all the data is

$$\mu = \frac{1}{|C|} \sum_{c \in C} \mu_c \tag{2-8}$$

The covariance within a class c is given by

$$\mathbf{V}_c = \frac{1}{|y_c| - 1} \sum_{i \in y_c} (y_c(i) - \mu_c)(y_c(i) - \mu_c)^{\mathbf{T}} \tag{2-9}$$

where we divide by $|y_c| - 1$ to obtain an unbiased estimate for the covariance. The average within-class covariance is equal to

$$\mathbf{W} = \frac{1}{|C|}\sum_{c \in C} \mathbf{V}_c \tag{2-10}$$

The between-class covariance is given by

$$\mathbf{B} = \frac{1}{|C|-1}\sum_{c \in C} (\mu_c - \mu)(\mu_c - \mu)^{\mathrm{T}} \tag{2-11}$$

We now wish to find set of orthogonal directions, z_i, that give a high variance between classes and a low within class variance. To achieve this we define the discriminant function

$$J(z) = \frac{z^{\mathrm{T}}\mathbf{B}z}{z^{\mathrm{T}}\mathbf{W}z} \tag{2-12}$$

The set of vectors which extremise $J(z)$ are found by setting its gradient to zero

$$\nabla J(z) = \frac{1}{z^{\mathrm{T}}\mathbf{W}z}(\mathbf{B}z - J(z)\mathbf{W}z) = 0 \tag{2-13}$$

The extremal vectors, z_i, therefore satisfy the generalised eigenvalue equation

$$\mathbf{B}z_i = \kappa_i \mathbf{W}z_i \tag{2-14}$$

where $\kappa_i = J(z_i)$. The solutions of the eigenvalue equation with the largest eigenvalues (i.e. the largest $J(z_i)$) are those that separate the classes most effectively. For a complete description on the practicalities of implementing canonical analysis see [45]. After the eigenvectors have been computed, the data is projected again into canonical space using k eigenvectors κ.

$$f_c(t) = [\kappa_1, \kappa_2, \ldots, \kappa_k]^T y_c(i) \tag{2-15}$$

By using the canonical eigenvectors, together with the principal component axes, we can project an image into the new space thus

$$f_c(t) = [\kappa_1, \kappa_2, \ldots, \kappa_k]^T [e_1, \ldots, e_k]^T x_c(i) \tag{2-16}$$

Classification is performed using the accumulated distance in canonical space. Each original image is now represented in canonical space by a low dimensional vector. Each sequence forms a trajectory in canonical space. The centroid of a reference sequence is the mean of the associated canonical projections. The accumulated distance is the sum of the distance from the reference centroid to each test image from a given sequence. The minimum distance from the test sequence to a centroid is used to label the subjects.

Huang et al [28] achieved high recognition rates of over 90%, but this was on a small database consisting of only six subjects. With a much larger database, the technical problems of implementing such a solution are manifold. The major problem is implementing principal component analysis as described above for larger databases where the size of the feature vector approaches the number of sequences in the database. There are two possible approaches to solving this problem:

- Incremental eigenanalysis
- Only using a subset of the images for PCA

An incremental eigenanalysis approach, such as [46], adjusts the eigenspace model when a new subject is added to the database. Essentially, this involves solving a smaller eigenproblem, which is used to rotate the original eigenspace model. This approach is still computationally intensive and thus may not be the best approach in this case. The second approach is to use a subset of the images to develop the eigenspace model. This is warranted in this case because the principal directions of variance do not vary much when a new subject is added to the database.

Principal component analysis combined with canonical analysis is a purely statistical approach to automatic gait recognition. It does not capture any temporal information about the subject's gait and effectively recognises a subject purely by variation in the static components of their silhouette. The classification procedure, using the accumulated distance in canonical space, loses dynamic information by only considering the centroid of the sequence in canonical space.

2.7 Summary and Aims of this Thesis

Medical and psychological studies have suggested that gait is a potential practical biometric. Automatic Gait recognition has had much attention devoted to it in recent times. Gait differs from traditional biometrics in that it consists of a sequence of images. This means that there is a tremendous amount of information to be exploited. Work by Collins et al [20], Huang et al[27], Hayfron-Acquah et al [25] and particularly Philips et al[31] have demonstrated that the silhouette information alone (gathered from the whole sequence) is enough for recognition. Cunado et al [22] and Yam et al [35] have developed model-based approaches which demonstrate that subjects can be recognised by dynamics of the manner in which they move. So far, statistical approaches have failed to devote enough attention to the dynamics of gait.

A new database has been developed to evaluate gait recognition algorithms. The new SOTON database is currently the largest of its kind consisting of 114 subjects filmed under laboratory conditions. Each subject consists of at least eight sequences, with a minimum of four walking left to right and four walking right to left. Subjects are filmed walking normal to the camera's plane of view. The subjects are filmed in front of a green background and chroma-key techniques are used to extract the human silhouette. This results in very high quality silhouettes that allow us to assess the basic properties of gait as a biometric. The silhouettes are further processed so that we have exactly one gait cycle of each person, starting from a known heel-strike.

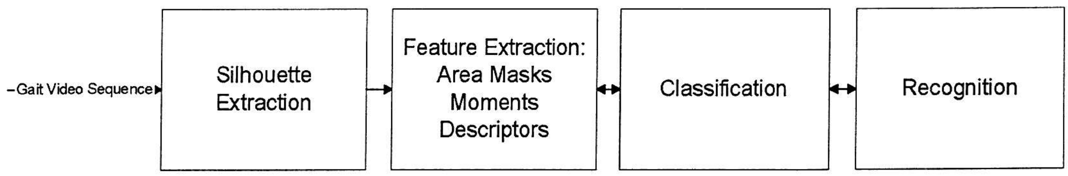


Figure 2 System Block Diagram

In this thesis, we propose two new approaches to automatic gait recognition. Figure 2 shows a block diagram of a generic approach to gait recognition. From an incoming gait video sequence, the silhouette is extracted. From this collection of silhouettes a feature vector is derived which describes the sequence. After generating a set of feature vectors for each sequence in the database, classification can take place and recognition rates can be calculated. Our two approaches to automatic gait recognition are area masks and moment descriptors, both of which generate feature vectors that describe a gait sequence.

The first of these two approaches, area masks, focuses on the temporal element in gait, which is overlooked in many statistical approaches. Area masks aim to extract a dynamic signal from the subject that represents the area change in portions of image sequence defined by area masks. This dynamic signal forms a feature vector that describes a gait sequence in compact form. By combining the information together from multiple area masks, a more descriptive feature vector can be formed with corresponding increases in recognition performance. The chief point of interest of this new approach is the use of temporal information as well as static silhouette information. In addition, the approach has low computational complexity, which is of great importance when typical gait databases consist of many gigabytes of data.

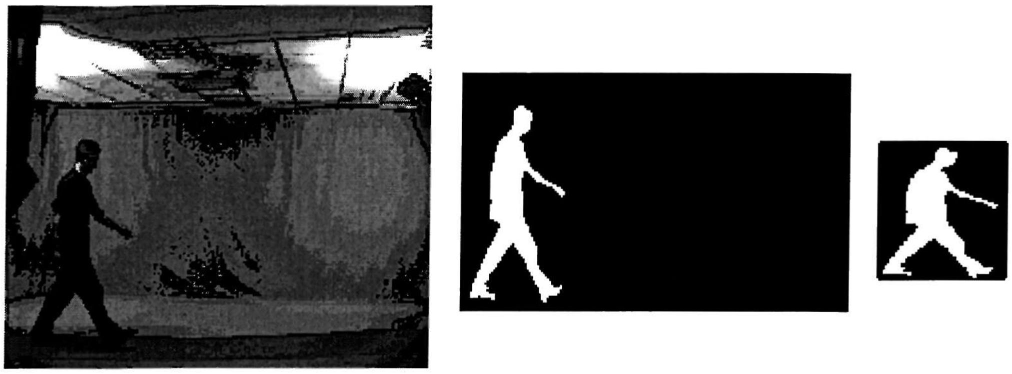
The second approach uses moment descriptors to describe a gait sequence by describing geometric properties in each frame of the sequence. Hu moments [50] are used, giving invariance to rotation, translation and scaling. To improve the quality of description we use a family of moment descriptors. A masking circle is used to define portions of the image to be described by moment invariants. The masking circle is proportional to the size of the image and is overlaid over the centre of mass of the silhouette, and we use moments to describe this new shape. This collection of moment descriptors describes the silhouette in more details and can hence improve the recognition capabilities.

Chapter 3 Gait recognition via Area Masks

3.1 Introduction

Gait differs from traditional biometrics in that it is both a function of space and time. The amount of information in a gait sequence is thus tremendous, and provides more information than single images. Approaches have been developed to take advantage of this [18, 25, 27] and they consider gait as a collection of silhouettes rather than a sequence. Area masks are a new technique for gait recognition that derive a dynamic signature of the subject, taking into account both the temporal and static components of gait. This dynamic signature represents area varying over time for specific regions of the image sequence defined by an area mask.

Before a dynamic signature can be obtained, the silhouette of the subject must be separated from the background. In part of the SOTON database, the gait sequences are derived in the laboratory, aiming for near-perfect conditions to obtain the best possible silhouette. Subjects are viewed fronto-parallel walking along a track in either direction. The camcorder's digital video is digitised via a firewire interface and eventually turn into individual frames. Figure 3(a) shows the original image from which the silhouette is extracted using chroma-key subtraction in conjunction with a connected components algorithm. The silhouette is then windowed to remove those parts of the background away from the chroma-key, such as the roof, Figure 3(b), and then resized to a 64×64 image, Figure 3(c). The image is resized in such a manner that each subject is the same height, but width information is preserved. The image size is set to 64 x 64 to reduce the computational cost. Examples of silhouettes of different subjects are shown in Figure 4.



a) Original Image b) Silhouette c) Windowed silhouette

Figure 3 Silhouette Extraction Process

Binary silhouettes remove information about colouring and interior shape but give invariance to clothing colour and most lighting conditions. Furthermore, by normalising the size of the silhouettes we remove variation due to distance from the camera, although we lose information about the subjects' heights. This invariance has the unfortunate side effect in that we have removed a known gender discriminator.

Figure 4 Sample Windowed Silhouettes

Following this process, we are left with a sequence of normalised and centralised silhouettes for each subject. Appendix A gives an overview of the entire SOTON database with stills from every subject and a complete sequence. We use multiple sequences each containing a single walking cycle (two steps) to perform recognition. This typically comprises of approximately thirty frames when filmed at twenty-five frames per second. We denote the images in the sequence by a vector $\mathbf{x}(t)$ where t labels the time when the image was taken.

In this chapter, we present area masks as a new technique for automatic gait recognition. Section 3.2 demonstrates how a dynamic signature is extracted from a gait sequence. Section 3.3 shows how performance can be increased by utilising gait's many degrees of freedom. Section 3.4 discusses the invariant properties of area masks and the concessions that have to be made when invariant properties are used. Section 3.5 examines the effects of extracting a dynamic signature when the subject is walking at an angle to the camera, and how this can be compensated by adjusting the size of the area mask in the simple case of the vertical line mask.

3.2 Extracting a Dynamic Signature

To obtain the dynamic signature we simply compute the area of the silhouette within a mask. These masks are chosen intuitively to isolate a particular part of the image. Examples of area masks are shown in Figure 5. Here, a horizontal line, Figure 5(a), isolates those parts in the region of the waist whereas a vertical one, Figure 5(b), selects the thorax and these parts of the legs intersecting with the vertical window. Many alternatives are possible such as a combination of the upper body and legs. Figure 5(c) merely measures the entire area change of the image. Figure 5(d) simply isolates the legs of the subject. Figure 5(e) isolates the area swept out by the legs. Figure 5(f) aims to isolate the area of central portions of the subject including the hips. Figure 5(g) isolates the top half of the subject together with the legs.

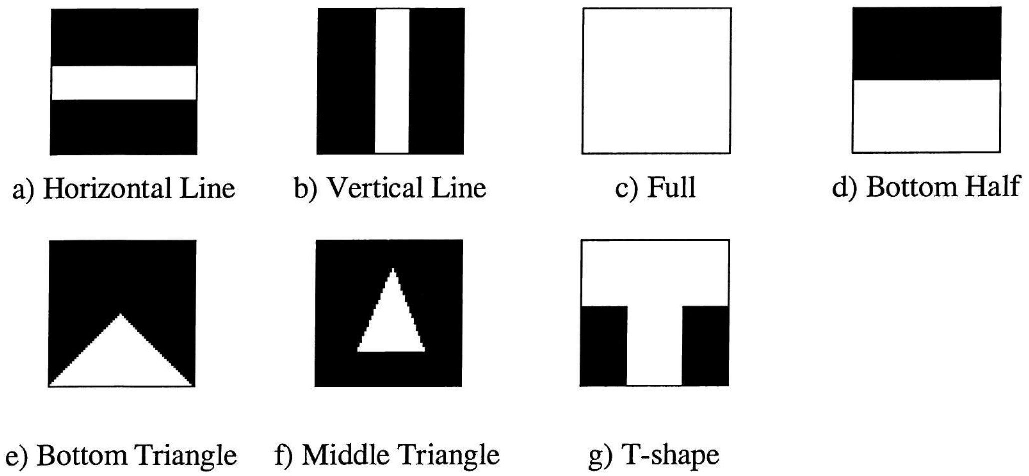


Figure 5 Sample Area Masks

Each mask, $\mathbf{m}_j$, represents an area mask, such as those shown in Figure 5, as a binary vector, with one representing the white parts of the image. Each image $\mathbf{x}(t)$ is also represented as a binary vector in the same manner. For each mask, we obtain a signature by determining the area between the mask and the images thus

$$S(t, j) = \mathbf{m}_j^{\mathrm{T}} \mathbf{x}(t) \tag{3-1}$$

We obtain an area history by applying each mask to each image in the sequence. Note that all the masks used are left right symmetric masks, and we therefore lose all information corresponding to the asymmetric part of gait. However, this also gives invariance to the direction of walk. That is, we could mirror any image within our sequence and we would obtain the same gait signature. We could use asymmetric masks, but we then need to reflect all walkers to walk in the same direction.

Examples of the gait sequences $S(t, j)$ for three different masks are shown in Figure 6. As can be seen, the dynamic signature (from some masks) is intimately related to gait. In the case of the bottom half mask, the peaks in the graph represent when the selected area of the subject is at a maximum (i.e. when the legs are furthest apart) and the troughs represent when the selected area is at a minimum (i.e. when the legs are closest together). The horizontal line mask selects those points near the region of the waist which varies less with thorax motion than the associated with movement of the subject's arms. Clearly, the masks emphasise bilateral symmetry: one leg swings whilst the other is in stance after which the limbs swap in function. Limping is an inherently asymmetrical gait, and this would be evidenced in clear disparity in the measured area between the two halves of the gait cycle.

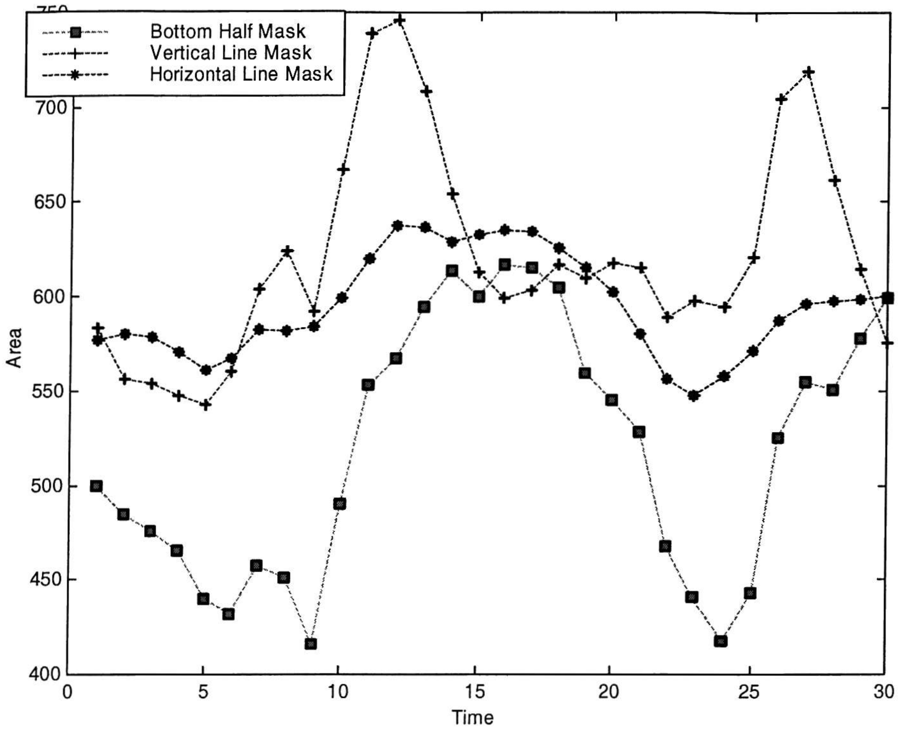


Figure 6 Sample Output for three Different Area Masks from a Single Subject

To compare sequences that start at different points in the gait cycle we have to align the sequences. We can do this by determining the minimum in a sequence and using this as a starting point. To eliminate variations due to the sampling time and speed at which the subject walks we select a part of the sequence corresponding to the full walking cycle and resample using cubic splines to interpolate between the observations $S(t, j)$. A cubic spline curve [47] is fitted for the whole gait cycle, and thirty evenly spaced samples are taken from the whole curve, giving a thirty element vector for each area mask used. Note that this does lose information about the subjects' speed. BenAbdelkader [17] has shown that this information can be used for recognition; however our technique is aimed at extracting the dynamics of area change and thus the speed of the walker can be removed. Figure 7 illustrates the difference between the feature vectors of two subjects walking in the same direction using the bottom half mask.

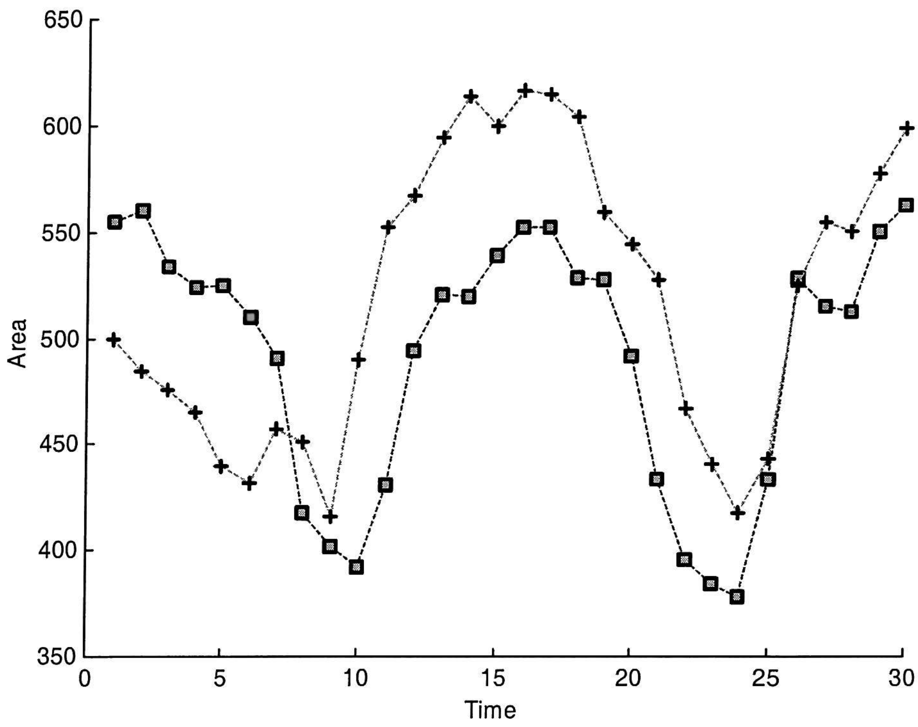


Figure 7 The Difference between two Subjects using the Bottom Half Mask

This feature vector is used to discriminate between subjects. Figure 7 clearly illustrates the differences between two subjects walking in the same direction. Both resultant vectors have similar sinusoidal shape but clearly differ in magnitude and relative timing allowing discrimination between the two subjects.

3.3 Combining Area Masks to Improve Performance

Each mask yields a vector describing the dynamics of area change within that mask. However, there are many degrees of freedom in a walking body, so we would expect that there would be a considerable degree of independent information from different masks. Therefore, we have combined the information from multiple masks to provide a more complete dynamics signature. We have chosen the simplest way to do this, namely to concatenate the thirty element vector from each area mask to form a single vector of $n \times 30$, where n denotes the number of area masks combined. We denote this dynamic signature by $\mathbf{v}$. We will have a different vector $\mathbf{v}$ for each sequence of a subject walking. We use these vectors to perform the recognition of subjects in our database.

3.4 Invariance Properties

Binary silhouettes are a necessary step in reducing the dimensionality of the data. By using binary silhouettes, we remove information about colouring and interior shape, but in return we get invariance to clothing colour and most lighting conditions. This is a good trade off as clothing and lighting conditions contain no gait information whatsoever. By normalising the size of the silhouettes, we also remove variation due to distance from the camera. By doing this we lose information about the subjects' height which is a known gender discriminator. This is unfortunate, but our aim is to recognise subjects by dynamic cues rather than static components such as body shape.

A further invariance area masks offer is a symmetrical invariance. We use symmetric area masks to isolate portions of the image, and this means that if the image is flipped about the y-axis then the area within the mask stays constant. This is a great advantage because it means we can treat subjects walking left to right with the same masks as those walking right to left.

3.5 Walking at an Angle to the Camera

The area masks essentially measure the area within the mask at a particular time t. What happens when the subject is walking oblique to the camera's view?

Let us consider the case of a simple pendulum operating under simple harmonic motion for the case of the vertical line mask. The pendulum has initial displacement x_0 and is of length l and we assume that it is infinitely thin. By assuming the line is infinitely thin, the area of the line is proportional to the length of the line. Let the angular velocity of the pendulum be ϖ which is constant.

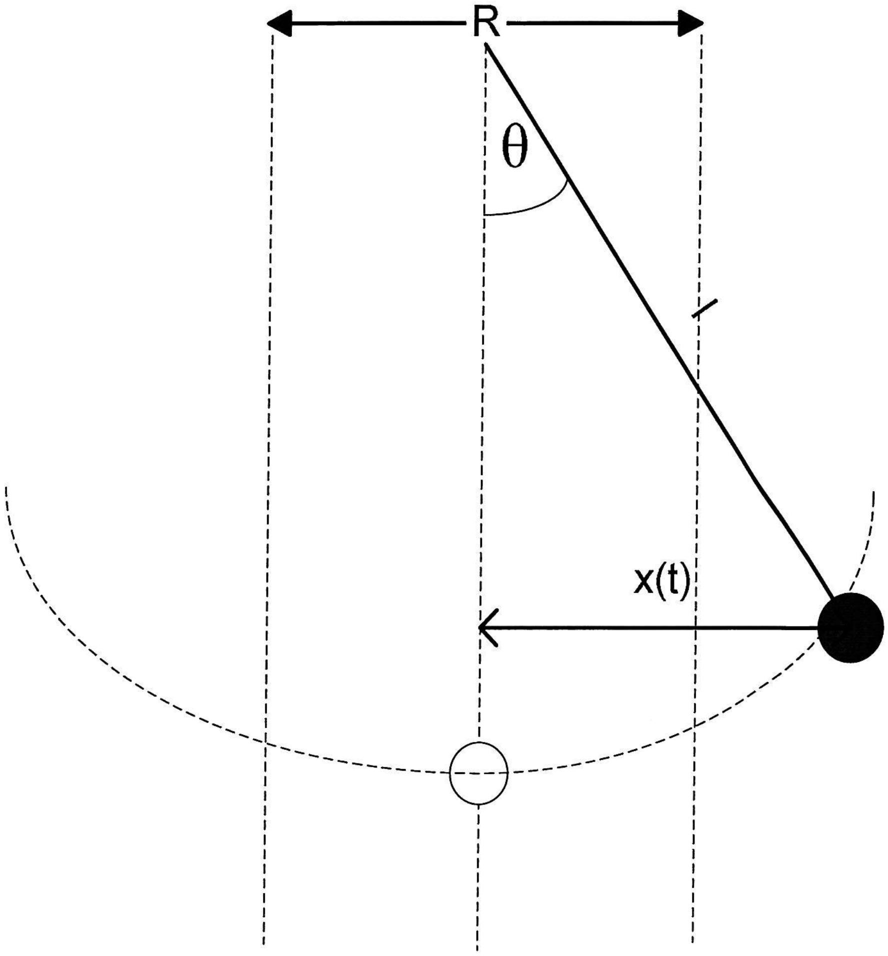


Figure 8 Simple Pendulum

R is a constant that defines the width of the vertical line mask. Let the origin of the system be at the fulcrum of the pendulum and the initial displacement of the pendulum be x_0. We can describe the current position of the pendulum bob x at time t by

$$x(t) = x_0 \cos(\varpi t) \tag{3-2}$$

$\theta(t)$ is the angle of the pendulum at time t, this is equal to

$$\theta(t) = \sin^{-1}\left(\frac{x(t)}{l}\right) \tag{3-3}$$

Since the line is infinitely thin, the length will be proportional to its area. Using simple trigonometry, we can calculate the length of the line inside the vertical line mask at time t as

$$A(t) = \frac{R}{2\sin\theta(t)} \tag{3-4}$$

However, if the magnitude of the position of the bob is less than $\frac{R}{2}$, then the entire length of the pendulum must be inside the area mask therefore $A(t) = l$.

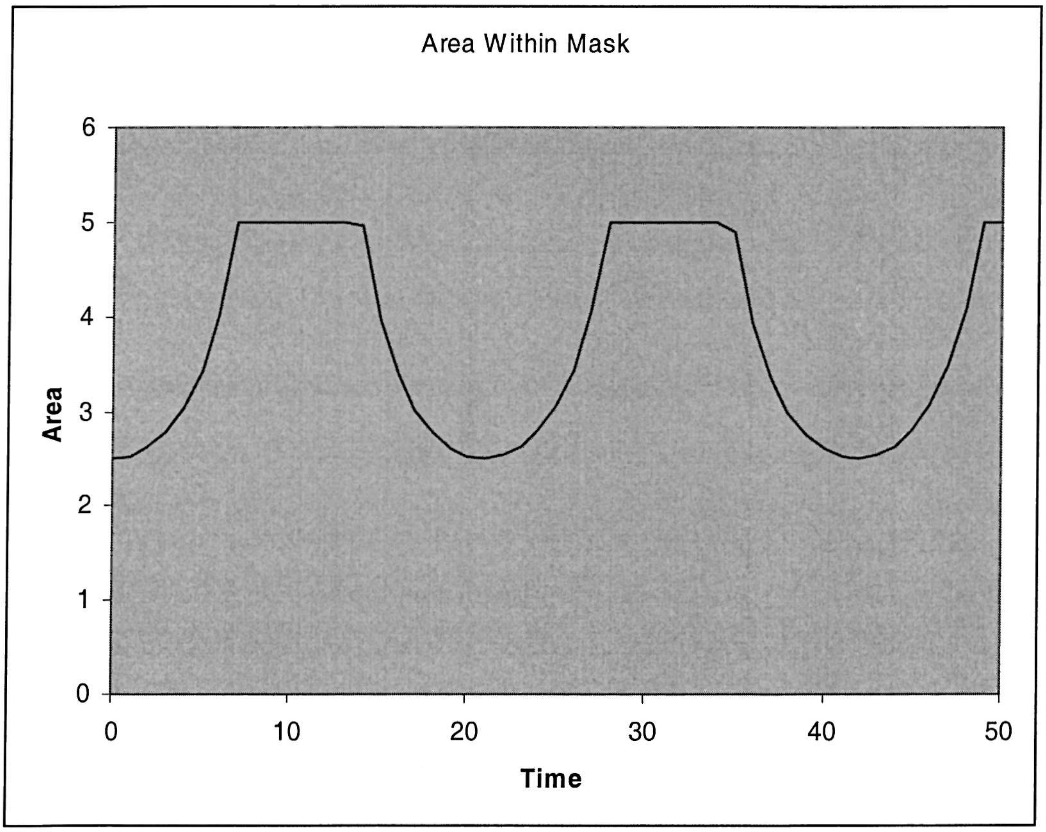


Figure 9 Area within Mask for Sample Pendulum

Figure 9 illustrates how the area changes over time for a sample pendulum. As can be seen the graph is sinusoidal in nature. The flat regions at the top of the graph occur when the entire length of the pendulum is within the area mask. The width of the flat regions at the top of the graph is proportional to the width of the gait mask R.

Suppose now that the pendulum is rotated about the y-axis by α degrees. This means that the position of the bob at time t will now be

$$x_\alpha(t) = x_0 \cos(\varpi t)\cos(\alpha) \tag{3-5}$$

So now, if the magnitude of the position of the bob multiplied by $\cos(\alpha)$ is less than $\frac{R}{2}$ then the entire length of the pendulum is inside the area mask, therefore $A(t) = l$ otherwise

$$A_\alpha(t) = \frac{R}{2\left(\frac{x(t)\cos(\alpha)}{l}\right)} \tag{3-6}$$

As the angle α increases, less of the motion of the pendulum is evident. When α approaches 90 degrees, the pendulum is almost parallel to the camera, hence no motion is visible from that viewpoint as the entire length of the pendulum is constantly within the area mask.

Figure 10 shows how the mean square error between $A(t)$ and $A_\alpha(t)$ varies for various angles α. As can be seen, as the angle of rotation increases, the mean square error increases exponentially.

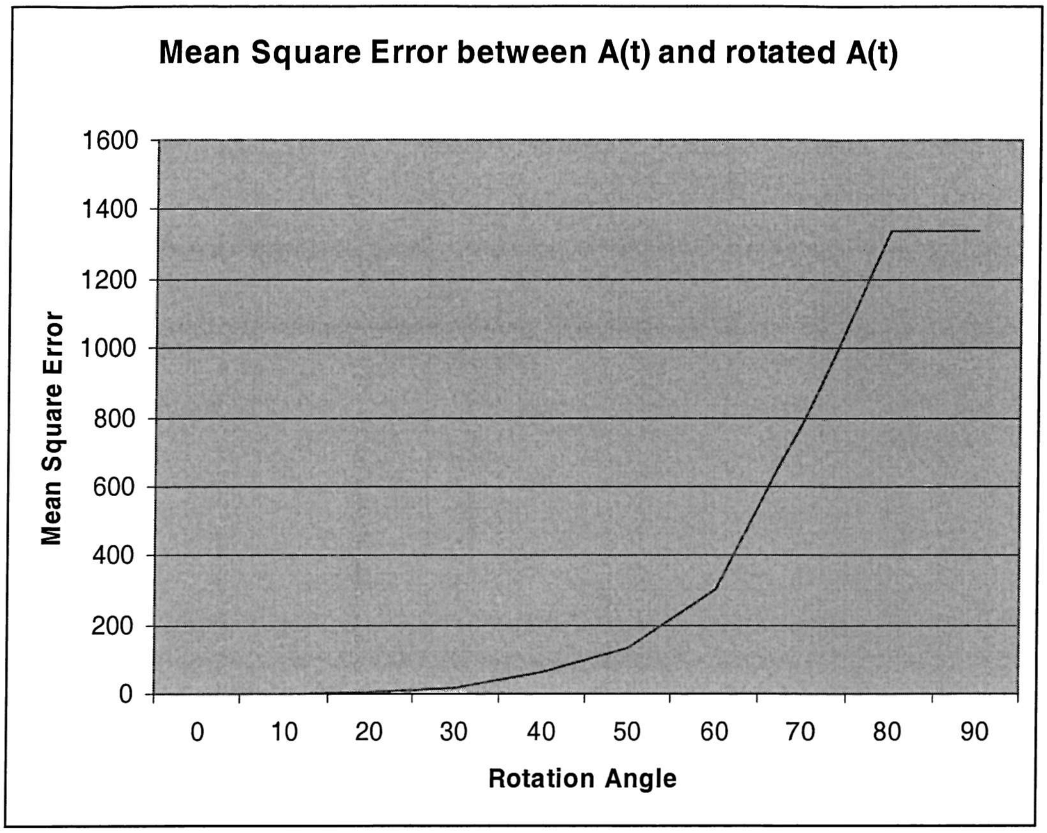


Figure 10 Mean Square Error between $A_0(t)$ and $A_\alpha(t)$

As can be seen from Figure 11, as the angle of rotation increases, the amount of information perceivable decreases. Indeed, once the angle of rotation is high enough, the entire length of the pendulum is constantly within the mask and thus no information whatsoever is obtainable.

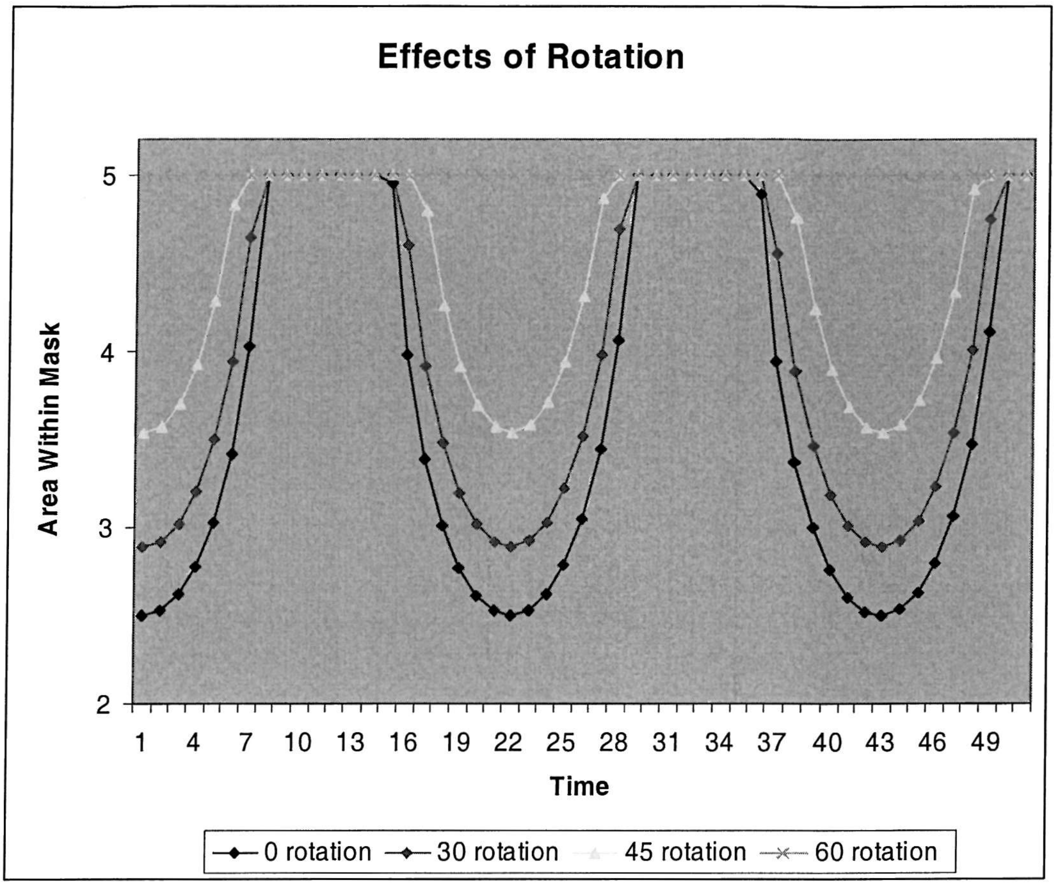


Figure 11 Effects of Rotation upon a Sample Pendulum

Obviously, when the pendular motion is parallel (or close to parallel) to the camera the area masks become useless since no pendular motion is perceptible at those angles, but how can we compensate for small rotations? By adjusting the width of the area mask, we are able to compensate for the change in angle. As the angle of the rotation changes, we need to adjust the size of the vertical line mask thus:

$$R_{\alpha} = R_0 \cos(\alpha) \quad \textbf{(3-7)}$$

Note that this is only an approximation, since at high values of α, the amount of information is insufficient for full compensation to be possible.

Angle of Rotation, α	Mean Square error between $A_0(t)$ and $A_\alpha(t)$ (with compensation factor $R_\alpha = R_0 \cos(\alpha)$)
0	0
10	0.02
20	0.12
30	0.33
40	0.63
50	0.99
60	1.35
70	1.66
80	1.87
90	1.94

Table 1 Effects of Rotation with a Compensation Factor for a Sample Pendulum

As can be seen from Table 1, by using the correction factor the mean square error is greatly reduced. Presently, no gait data is available with a subject not walking normal to the camera plane of view so no empirical results can be obtained to support the hypothesis that small angles of rotation can be compensated for. This work suggests that it is possible to compensate for small rotational angles, if the angle of rotation is known. This relates to work by Carter and Nixon [48] who examined the effects of trajectory on gait signatures.

3.6 Summary

This chapter has presented a new approach to automatic gait recognition that focuses on the temporal element of gait. A dynamic signature is formed by determining the area change in specific portions of the image defined by the area masks. Area masks are chosen intuitively to isolate specific regions of the gait sequence. The technique is performed upon a single gait cycle (heel strike to heel strike) and as such requires the identification of key frames in the image sequence. These key frames can be located by finding minima in an area signature or by alternative techniques described by Collins et al [20]. Following the identification of the key frames to be used, the area is calculated for the sequence using various portions of the image (defined by the area masks). This results in a vector for each mask used that describes the area change for that particular portion of the image sequence. These vectors will be of different lengths for different subjects, since each subject may take a different amount of time to complete a single gait cycle. To compensate for this, a spline curve is fitted to the data and thirty samples are taken from each curve, resulting in a thirty-element vector for each mask used.

Each mask results in a thirty element vector describing the sequence in compact form. By combining these feature vectors, we can take advantage of gaits' many degrees of freedom and increase the recognition performance. Vectors are combined by simple concatenation resulting in a vector of size n by 30, where n is the number of area masks combined.

The masks chosen are left right symmetric which means that the area description is invariant to whether the subject is walking left right or vice versa. By normalising so that only one gait cycle is present, using area masks is also invariant to the speed of walk, assuming that the subjects walk does not change as their speed varies. As part of the silhouette extraction process, subjects are normalised for height, which has the unfortunate side effect of removing a known gender discriminator. Using binary silhouettes discards colouring and interior shape information, but does give invariance to clothing colour and most lighting conditions. In addition, we have shown the possibility of compensating for the subject walking at small angles to the camera, though presently no database exists for us to test this hypothesis empirically.

In the next chapter, we introduce another new technique for automatic gait recognition, moment based descriptors. This approach aims to describe, in terms of geometric invariants, the silhouettes in each gait sequence. By using a family of invariant descriptors the quality of the shape description can be greatly improved, in much the same way that combining area masks yields increased performance.

Chapter 4 Gait Recognition by Moment Based Descriptors

4.1 Introduction

Previously in Chapter 3, we have used area masks to describe the dynamics of area change through a gait sequence and this has formed the basis of description. Moments are an alternative approach for describing a shape and they can provide much more information than a single measure of area. Lower order moments provide information similar to area masks, such as total area of the shape. We use moments to describe information for each image in the gait sequence and this forms our basis of description. By using Hu invariant moments [50], the technique is invariant to rotation, translation and scaling of the silhouette.

A gait sequence is once again represented as a collection of binary silhouettes. The binary silhouettes are generated using the same procedure described in section 3.1. These binary silhouettes provide us with invariance to most lighting conditions and clothing colour, but do lose some information. As an alternative, it would have been possible to use images such as those shown in Figure 3 (b), as the Hu moments used to generate the feature vectors are invariant to translation and scaling. The computational cost of doing so on a large image is prohibitive, so windowed silhouettes were used.

In this chapter, we explain why we have used moments to describe a gait sequence. Section 4.2 describes how we can use a family of moment descriptors to describe a single shape. This relates to work from Sluzek [49] and forms the basis of our recognition from sequences. Section 4.3 describes how Sluzek's approach has been extended to describe a sequence of images. Finally, section 4.4 discusses the invariance properties of using this approach.

4.2 Moment Based Descriptors for a Single Image

Moments are calculated on a bounded region R in the XY plane. Moments of order p, q of R are defined as:

$$m_{pq} = \iint_R x^p y^q dxdy \tag{4-1}$$

For discrete images, this becomes:

$$m_{pq} = \sum_{(i,j)\in R} i^p j^q \tag{4-2}$$

Centralised moments are defined with the co-ordinate system translated to the centre of R, (a,b).

$$M_{pq} = \iint_R (x-a)^p (y-b)^q dxdy \tag{4-3}$$

For discrete images this becomes

$$M_{pq} = \sum_{i,j\in R} (i-a)^p (j-b)^q \tag{4-4}$$

a and b are terms to centralise the moments according to centres of mass and are defined thus

$$a = \frac{m_{10}}{m_{00}} \qquad b = \frac{m_{01}}{m_{00}} \tag{4-5}$$

The lower order moments are simple properties used to describe the shape. For binary shapes, m_{00} describes the shape in terms of its total area. m_{10} represents the mean x co-ordinate of the shape, and similarly m_{01} represents the mean y co-ordinate of the shape.

Hu has derived moment expressions that are invariant to translation, rotation and scaling of shapes. Some examples are given below. The moments omitted, I_3 to I_7, are usually assigned to moment invariants of order 3 [50], but we are only using lower order moment descriptors.

$$I_1 = \frac{M_{20} + M_{02}}{(m_{00})^2} \tag{4-6}$$

$$I_2 = \frac{(M_{20} - M_{02})^2 + 4M_{11}^2}{(m_{00})^4}$$

$$I_8 = \frac{M_{20}M_{02} - M_{11}^2}{(m_{00})^4}$$

I_8 is invariant to any non-singular linear mapping of R [51]. For digital images, these moments are only approximately invariant to translation, rotation and scaling because of quantisation errors. These moments can only be used to differentiate between images of real objects if they are significantly different. More features are needed to discriminate between similar items. Using higher order moments is not a solution as higher order moments are more sensitive to minor shape deformations. For a given shape R and selected invariant I_k the magnitude of the error $V(R, I_k)$ (due to quantisation errors, camera non-linearity) can be estimated from a sufficient number of images (captured in diversified conditions) of R by simply comparing theoretical values with those calculated empirically. It is possible to evaluate whether the invariant I_k can differentiate between two shapes R_1 and R_2 using the following discrimination parameter:

$$Q(I_k, R_1, R_2) = \frac{V(R_1, I_k) + V(R_2, I_k)}{|I_k(R_1) - I_k(R_2)|} \tag{4-7}$$

where $I_k(R_1)$ and $I_k(R_2)$ are the theoretical values of the invariant for R_1 and R_2. If $Q(I_k, R_1, R_2)$ exceeds one, the invariant I_k cannot discriminate between the two images. This condition can occur frequently, especially with higher order moments.

Sluzek [49] proposed a method to overcome these difficulties by using a family of shape descriptors. By using additional descriptors, it is possible to improve the quality of descriptions. Let C_α be the circle defined as

$$C_\alpha = (x-a)^2 + (y-b)^2 = \frac{\alpha m_{00}}{\pi} \tag{4-8}$$

The area of the circle is equal to the area of the shape multiplied by α. The value of $\alpha_{\max}$ is selected so that the circle does not occlude the entire area of R and always a part of the shape R is visible. Let us denote the region R which is not occluded by the circle as $R(\alpha)$ which is defined as

$$R(\alpha) = \left\{ (x,t) \in R \wedge \left|(x,y) - (a,b)\right|^2 > \frac{\alpha m_{00}}{\pi} \right\} \tag{4-9}$$

Thus for a given shape R (i.e. such as those shown in Figure 4) we can create functions $I_k(\alpha)$ for each moment k which operates on the region $R(\alpha)$. These functions depend on both the region R and the value of α and are invariant to rotation, translation and scaling and are known as *m-invariant* functions. The function $I_k(\alpha)$ is the k^{th} Hu invariant on the region $R(\alpha)$. Note that when calculating $I_k(\alpha)$, a and b must be recalculated on the new region $R(\alpha)$.

A given shape/image can now be represented by a family of shape descriptors. Adequate sampling should be chosen to give a detailed representation of the shape. $I_k(\alpha)$ is then the signature of the image for a given range of circles, α. Figure 12 shows a range of moment descriptors for a single gait image. In this case, the first Hu invariant moment was used with α ranging from zero to 2.0 in steps of 0.2.

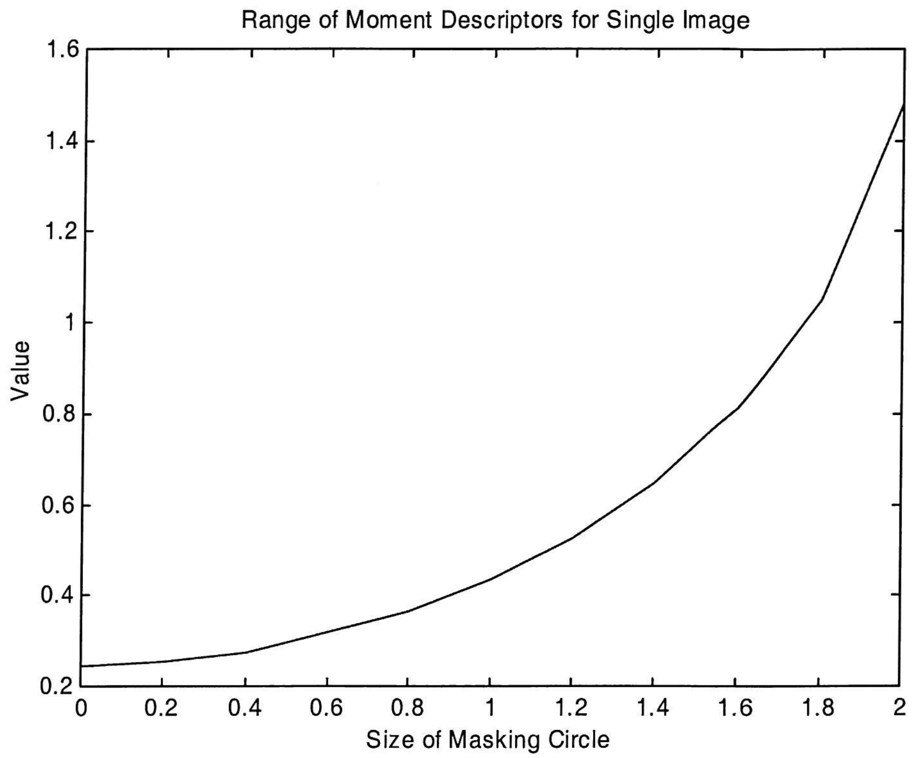


Figure 12 Family of Moment Descriptors for Single Image

These moment descriptors can be used to discriminate between two different images. Sluzek [49] showed that these shape descriptors are more reliable than traditional moment shape descriptors for identifying undamaged objects and/or to detect damaged ones. Sluzek ran an experiment on a database of 130 images and achieved a hundred percent accuracy both in the identification of objects and in the quality inspection.

Figure 13 shows the moment descriptors for two single images from the walking sequence of different subjects. As can be seen from the graph, there is little visible difference between the two graphs when the value of α is low, but as the value of α rises the differences between the two becomes more apparent. Gait consists of a temporal, as well as spatial component, so over a sequence the differences between subjects should become even more apparent as the small differences will accumulate over an entire sequence.

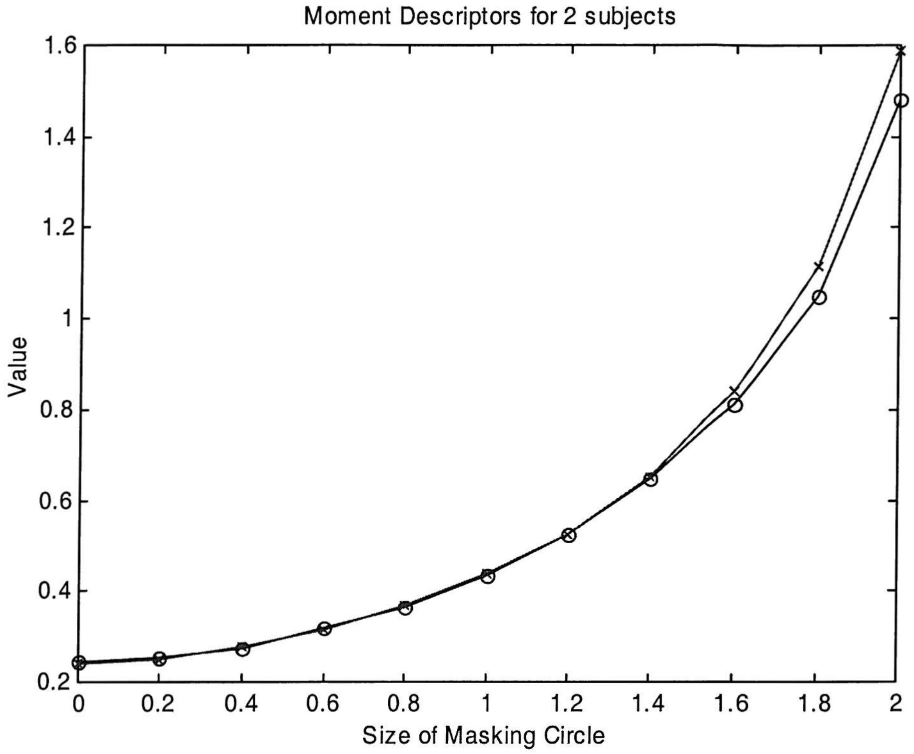


Figure 13 Moment Descriptors for Two Subjects

In the following section we discuss how we have applied this technique to automatic gait recognition, by extending the approach to deal with sequences of images.

4.3 Recognition of Gait using Moment Based Descriptors

Gait recognition differs from traditional biometrics such as iris and fingerprint because it requires identification of a subject from a *sequence* of images. A gait sequence consists of a large amount of information, both spatial and temporal, and thus a technique is needed to represent the information in compact form. Let us consider a gait sequence as a collection of regions denoted by R_t where *t* labels the position of the region in the gait sequence. For each individual image R_t, we can once again define regions masked by circles as $R_t(\alpha)$. As the size of the masking circle increases, we describe areas on the periphery of the image, such as the subject's feet. These are the areas of the image most likely to vary considerably over a gait sequence.

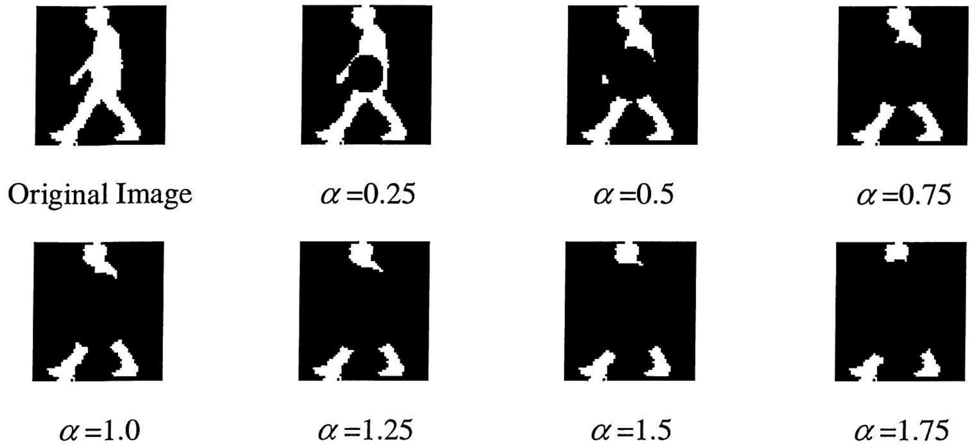


Figure 14 Masking Circles for various values of α

Figure 14 illustrates how the masking circle affects an image for various values of α. We can now define a complete signature set for a gait sequence thus

$$S = \{I_k(t,\alpha) \mid k = 1,2,8; t = 0,N; \alpha = \alpha_{\min}, \alpha_{\min} + \alpha_{step}, ..., \alpha_{\max}\} \tag{4-10}$$

t is chosen such that a full gait cycle was sampled at intervals of $\frac{1}{N}$ starting from a known start point. In the case of the SOTON database, all subjects start from a heel-strike and exactly one gait cycle is present. In other databases, key frames could be located using techniques developed by Collins et al [20]. A spline curve was used to interpolate between values to provide exactly thirty samples for a full gait cycle. Each sequence can therefore be described as a set as above for values of k and α. Figure 15 shows the sample output for a single subject using m-invariant functions for each frame in the sequence. The Hu1, Hu2 and Hu8 moments are illustrated and the size of the masking circle is zero (i.e. no masking circle is used).

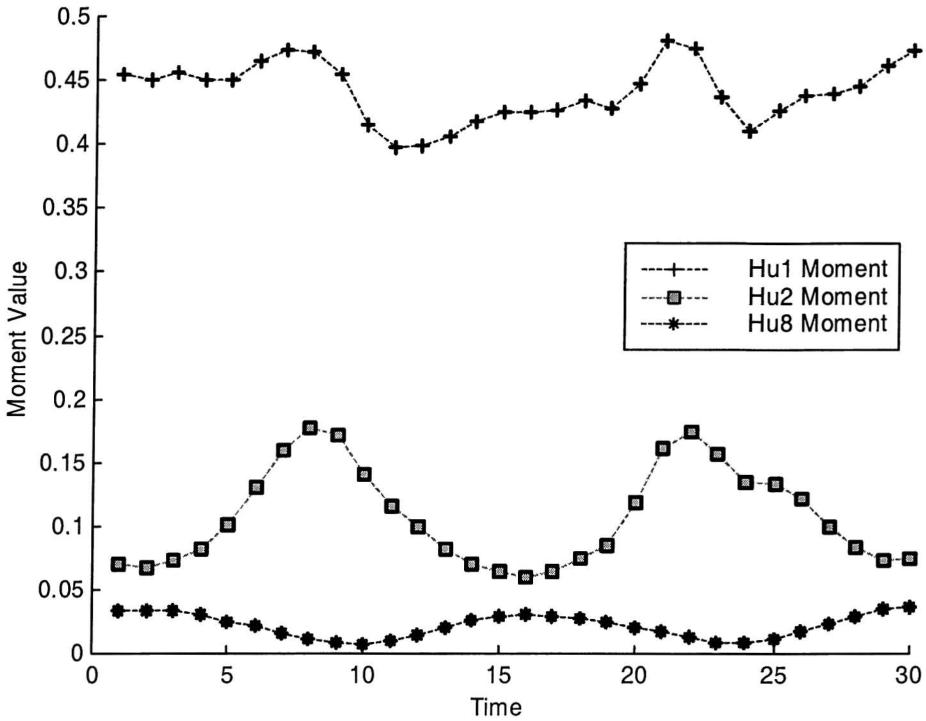


Figure 15 Hu Moments for a Subject with α =0

A suitable α range must be chosen to give adequate image representation. In practice, if the value of α is too large then no information about the image is obtained and this can lead to poor recognition performance. To ensure that all sequences are the same length we use a cubic spline curve to represent the data and take the same amount of samples (thirty) from each sequence $S_{\alpha,k}(t)$. Figure 16 illustrates a spline curve fitted to a plot of the Hu2 moments with α equal to zero.

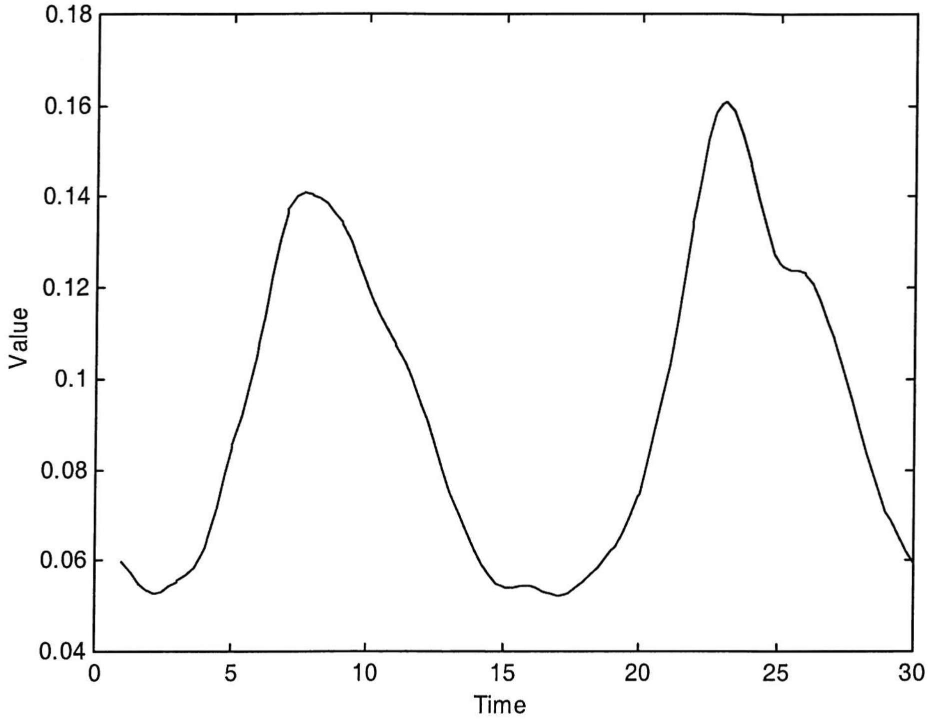


Figure 16 Hu2 Moments for a Subject with α =0 with a Cubic Spline Curve Fitted.

After sampling, each subject is represented by the same amount of data. Each value of α (the masking circle) and k (the Hu moment used) yield a vector describing how the sequence changes over time. These vectors are combined by concatenation to produce a single vector that describe the sequence. This complete sequence description for Hu1 moments is illustrated in Figure 17. This is the moment based description using Hu1 moments for a single gait sequence concatenated with moment based descriptors for the same sequence with increasing alpha values. The upward trend in the graph corresponds to the general increase of moment values when larger values of alpha are used.

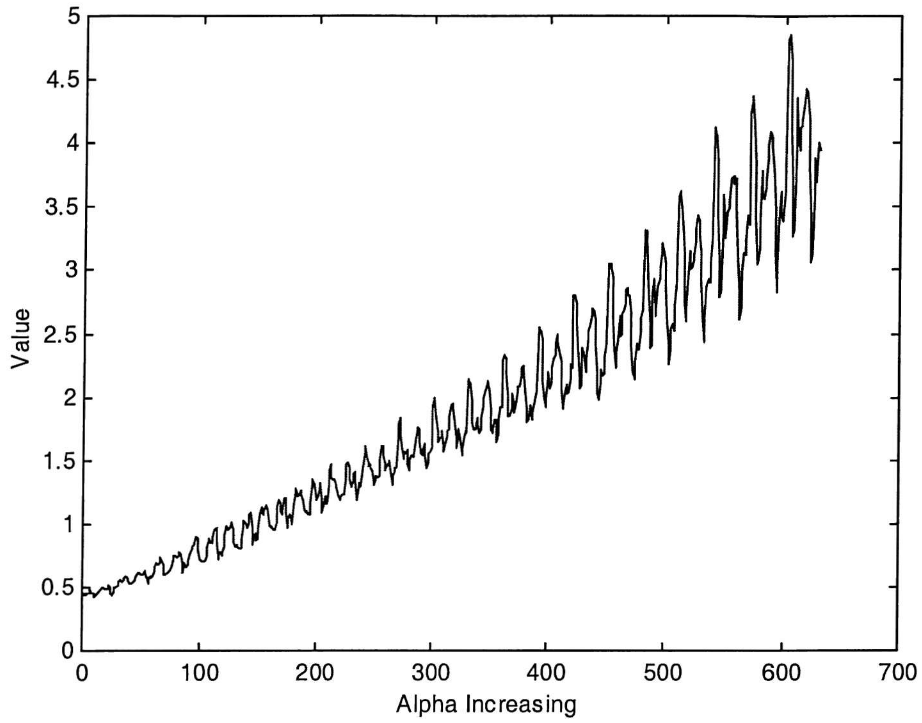


Figure 17 Hu1 Moments for a Subject with α =0 to α =2 in Steps of 0.1

For each sequence in the database, a complete sequence description (for each moment, Hu1, Hu2 and Hu8) is calculated. This description is our feature vector that we use for the purposes of classification.

4.4 Invariance Properties

Invariance properties arising from using binary windowed silhouettes have been previously discussed in section 3.4. They include invariance to the following:

- Distance to the camera
- Height
- Clothing Colour
- Lighting changes

As a consequence of using Hu invariant moments, the moment descriptors for each silhouette are invariant to rotation, translation and scaling. Invariance to rotation could be of benefit if the subject is walking uphill or downhill, but this implies that gait stays constant when walking at an incline, which is certainly not true [52].

Scaling and translation invariance are important because they mean that silhouette sequences from different databases can be compared easily. Scaling invariance is of particular importance as gait can be filmed from a great distance and therefore the resolution of such images may be lower than those used in the SOTON database. In addition, scaling invariance gives us immunity to camera zoom. Translation invariance removes sensitivity to the centring of the silhouette, which is a problem when area masks are used.

4.5 Summary

Moment based descriptors are a new technique in automatic gait recognition that describes the gait sequence in terms of invariant properties for each image in the sequence. By using Hu invariant moments the description is invariant to rotation, translation and scaling of the original images. A family of shape descriptors is used to describe each silhouette, which aims to provide a more complete description of each image and thus better recognition performance.

The invariance properties of rotation and translation are of minimal advantage when using windowed silhouettes (since the subject is centred in the image and usually not rotated). However, scale invariance is very important as it means silhouettes recorded at different resolutions can be compared. Rotation invariance may be of importance when a subject is walking at an incline, however this relies on the assumption that a subjects gait does not change when they are walking at an incline. Rotation invariance also has the advantage that if the camera is misaligned then results can still be obtained.

For each image, a family of Hu invariant moment descriptors is calculated. A range of masking circles, whose size is proportional to the size of the images, are placed over the image, and the region outside the image is then described using Hu invariant moment descriptors. We form a family of descriptors for each image in the sequence, and then concatenate these to form a family of descriptors for the entire sequence.

We have assumed that each gait sequence we are presented with consisted of exactly one gait cycle, which is true for the case of the SOTON database. In other databases, where this may not be the case, one gait cycle can be extracted by identifying key frames in the image sequence as described by Collins et al. [20].

Since each sequence may be of varying length, we use a cubic spline curve to interpolate and make each description of uniform length. Classification is performed using the k nearest neighbour technique in feature space.

In the following chapters, we evaluate both algorithms on a variety of tasks. Both techniques are tested on the tasks of person discrimination, gender discrimination and performance analysis. In addition, we look at whether gait is symmetric and whether information about heel strikes and direction of travel can be used to improve the recognition results.

Chapter 5 Automatic Gait Recognition

5.1 Introduction

The section presents results using gait as a biometric on the new SOTON database. The part of the SOTON database used for the testing procedure consists of 114 subjects with at least eight samples of each subject. Subjects walk normal to the camera's plane of view and are filmed under laboratory conditions. By filming under controlled conditions, we can concentrate on the basic properties of gait as a biometric, rather than the problem of extracting clean data. The background of the image is removed using chroma-key techniques, and the image is cropped and resized to a 64 x 64 silhouette with the subject centred in the image. Examples of these images are shown in Figure 4, and a complete overview of the database is available in Appendix A

Section 5.2 discusses why we have chosen to use a simple *k* nearest neighbour classifier algorithm rather than a more sophisticated classifier such as canonical analysis. Section 5.3 looks at performing automatic gait recognition by using area masks. In that section, we examine recognition rates using a single area mask and then demonstrate how performance can be improved by combining the results from multiple masks. We also report on studies where we focus on recognising a subject purely from the temporal components rather than any static cues such as body shape. In addition, we try acquiring information from only the legs of a subject and determine whether this alone is enough to discriminate between subjects. Section 5.4 looks at automatic gait recognition by using moment based descriptors. We show how by using a family of shape descriptors the performance can be improved. In the final section, we discuss what can be learnt from the results from this chapter.

5.2 Classification

Both area masks and moment descriptors generate a feature vector that describes each sequence in the database. Many classification techniques are available, ranging from simple approaches such as *k* nearest neighbour algorithms, to complicated algorithms such as canonical analysis or support vector machines.

Canonical analysis has been used before for automatic gait recognition [28] but only upon a small database consisting of six subjects. For larger databases, canonical analysis has a high computational cost. In particular, when the size of the feature vector approaches the number of subjects in the database, the computational framework becomes unstable [45]. This means that an extra step needs to be added into the canonical analysis routine that reduces the size of the feature vector using principal component analysis. By incorporating principal component analysis into the classification procedure, the computational cost may reduce (since the feature vectors will reduce in size by the time canonical analysis is applied); however, the complexity of the technique increases.

A *k* nearest neighbour algorithm provides a basic classification method that is fast to compute and produces easily comprehensible results. Essentially, a *k* nearest neighbour algorithm measures the Euclidean distance between a test sample and the training samples in feature space. In a *k* nearest neighbour algorithm, a training set $\mathbf{T}$ is used to determine the class of a previously unseen sample $\mathbf{X}$. A Euclidean distance metric is used to find *k* closest training samples to our unknown $\mathbf{X}$. Of these *k* closest samples, whichever occurs most frequently is used to label our unseen sample $\mathbf{X}$. In practise a *k* nearest neighbour algorithm can be difficult to beat, even with sophisticated techniques. One of the primary advantages of a *k* nearest neighbour algorithms is that it is very fast to compute.

By using a *k* nearest neighbour algorithm and a simple Euclidian distance metric, we get an idea of the baseline performance of the system and more experiments can be undertaken to assess properties of each algorithm (area masks and moment descriptors). In each experiment we have used leave one out cross validation which provides a more robust measure of recognition performance [53].

Table 2 shows the effects of using a different classifier using a single area mask. The *k* nearest neighbour algorithm was applied in feature space and the value of *k* was one. Canonical analysis was applied in the following manner. Canonical analysis was used to project the training data into canonical space. The centroids of each class were calculated in canonical space, and the test sample was then projected into the same canonical space and the nearest class centroid was used to label the sample. The results in brackets for canonical analysis indicate the recognition performance when a *k* nearest neighbour algorithm, with *k*=1, is applied in canonical space (in place of labelling based on the nearest centroid). As can be seen from Table 2 k nearest neighbour offers much greater recognition performance in all circumstances.

Area Mask	**Recognition Rate**	
	Canonical Analysis	k Nearest Neighbour
Bottom Half Mask	15.8% (13.2%)	35.6%
Horizontal Line Mask	13.5% (11.7%)	41.3%
Vertical Line Mask	22.7% (14.1%)	43.4%

Table 2 Comparison of Classification Methods

The poor performance of the canonical analysis algorithm can be attributed to poor clustering in canonical space. The canonical analysis transform maps the training data into a space where intra-class variance is minimised whilst inter-class variance is maximised. The performance of canonical analysis on the training set is poor in this case and that is why the performance is so poor compared to the k nearest neighbour algorithm.

In our work, we have assumed gait to be a symmetrical pattern of movement. In Chapter 7, we examine this assumption in detail and show gait to have an asymmetrical component. By utilising this component of gait, we demonstrate how the performance of the canonical analysis classifier can be improved.

When we refer to using a k nearest neighbour algorithm we have chosen k to be equal to one unless mentioned otherwise, since this gave best performance in most circumstances. Table 3 illustrates how the performance varies for a sample of recognition algorithms using different values of k. As can be seen, as the value of k increases, the recognition rate declines. For this reason we have decided to use k=1 in all of the tests. In the case of moment based descriptors, we used the Hu1 moments with alpha ranging from 0 to 1 inclusive in steps of 0.1.

Algorithm	Recognition Rate		
	k=1	*k*=3	*k*=5
Bottom Half Mask	35.6%	20.7%	10.0%
Top Half Mask	28.2%	12.2%	5.6%
Vertical Line Mask	43.4%	27.1%	16.1%
Seven area masks combined	71.9%	54.5%	36.6%
Hu1 Moment Based Descriptor	68.3%	42.5%	25.1%

Table 3 Using a *k* Nearest Neighbour Algorithm

The performance degradation which arises when higher values of k are used suggests that the clustering of classes in Euclidean space is poor. In the tests we ran in Table 3, we used the entire SOTON database including eight samples of each subject (four filmed in each direction). In Chapter 7 we examine the symmetry of gait and show that by taking into account heel strikes and direction of travel the results can be greatly improved. This suggests one reason why the performance decreases when using high values of k for the k nearest neighbour algorithm.

More advanced classification methods could be employed, such as support vector machines, however that could have led us to focus solely on recognition performance rather than improving the quality of the underlying feature vector. For this reason, we have chosen a simple classifier as it allows us to perform many experiments and analyse the results in detail.

Throughout the following chapters, when we refer to classification we have used a k nearest neighbour algorithm (using the Euclidean distance metric with k=1) in feature space unless mentioned otherwise.

In conclusion, we have decided to use a k nearest neighbour algorithm for the following reasons

- Similar / Greater performance than using canonical analysis
- Lower computational complexity than canonical analysis, especially when using large feature vectors (as in the case of moment descriptors)

5.3 *Person Identification using Area Masks*

5.3.1 Using a Single Area Mask

Table 4 shows the results using a single area mask on the SOTON database. A *k* nearest neighbour algorithm was used with leave one out cross validation. Eight samples of each subject were used with four walking left to right and the remaining four walking right to left.

Area Mask	Recognition Rate
Bottom Half (BH)	35.6%
Bottom Triangle (BT)	38.2%
Horizontal Line (HL)	41.3%
Middle Triangle (MT)	39.4%
Top Half (TH)	28.2%
T-shape (TS)	43.2%
Vertical Line (VL)	43.4%

Table 4 Recognition Rates from Single Masks

Using a single mask, recognition rates range from just under 30% to over 43%. Whilst this performance may appear disappointing, the chance recognition rate is $\frac{1}{114}$, and the achieved recognition rates are significantly better than this. The lowest performance comes from the top half mask, which suggests that the area change in the top half of a subject's silhouette provides far less discrimination than the area change in the bottom half of a subject's silhouette. In the top half mask the only major area change is associated with arm swing which suggests that the movement of the arms alone does not offer discriminating capability.

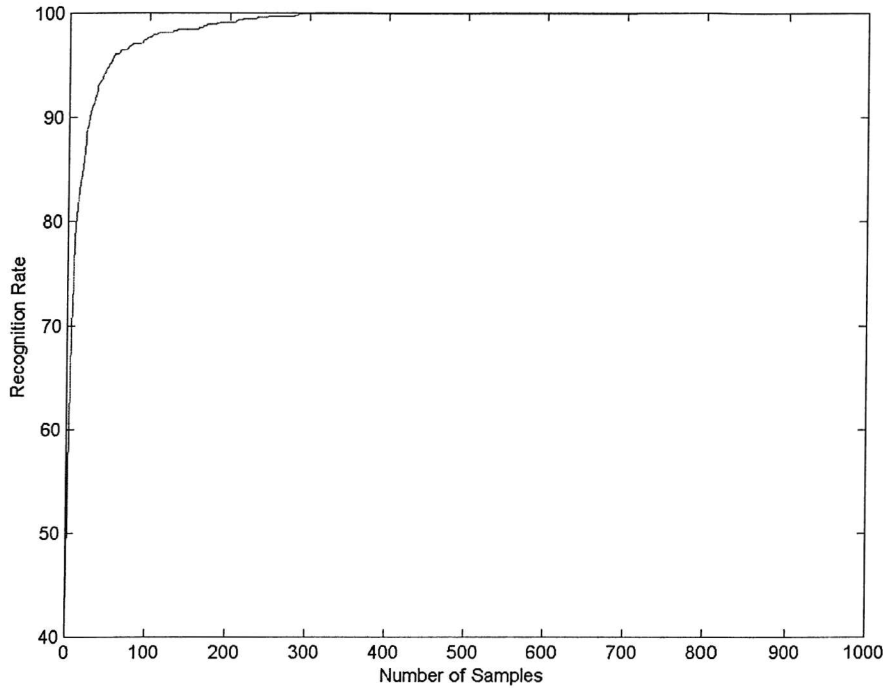


Figure 18 Cumulative Match Score Graph for Vertical Line Mask

Figure 18 shows the cumulative match score graph for the vertical line mask. As can be seen the initial recognition rate of just over 40% rises rapidly in the cumulative match graph. This indicates that the feature vector from just a single mask is a good discriminator, and more performance potential exists.

5.3.2 Improving Performance by Combining Masks

By combining the feature vectors from multiple masks together by concatenation, we hope to improve the recognition results. Table 5 shows the recognition results from combining three different selections of area masks.

Area Masks Combined	Recognition Results
Bottom Half + Top Half	58.0%
Horizontal Line Mask + Vertical Line Mask	59.2%
All Four Masks shown above	70.8%

Table 5 Results from Combining Multiple Masks

By combining just two area masks, the performance rate improves to nearly 60%. By combining the results from four area masks (BH, HL, TH, VL) the recognition result leaps again to nearly 71%.

Masks	BH	BT	HL	MT	TH	TS	VL
BH		45.3%	56.1%	51.0%	58.0%	59.2%	58.6%
BT	45.3%		61.2%	54.9%	61.5%	59.9%	58.6%
HL	56.1%	61.2%		59.9%	57.0%	59.9%	59.2%
MT	51.0%	54.9%	59.9%		61.2%	57.2%	57.1%
TH	58.0%	61.5%	57.0%	61.2%		58.9%	57.5%
TS	59.2%	59.9%	59.9%	57.2%	58.9%		43.8%
VL	58.6%	58.6%	59.2%	57.1%	57.5%	43.8%	

Table 6 Combining Pairs of Masks

Table 6 shows the recognition rate when each pair of masks is combined. The performance level for combining two masks varies from 43.8% to 61.2%. We hypothesize that the degree of independent information from each mask is related to the amount of overlap between the two masks (e.g. the more overlap between the masks, the less independent information). The worst performance in Table 6 comes from the vertical line and t-shape mask, which have a high degree of overlap and thus not much significantly different information. The best level of performance comes from the combination of the top half mask and the bottom triangle mask, which feature a small amount of overlap.

Figure 19 illustrates the relationship between the amount of overlapping pixels and the recognition rate when two area masks are used. A linear line of best fit demonstrates the trend that the more pixels masks have in common the worse the recognition performance for pairs of area masks.

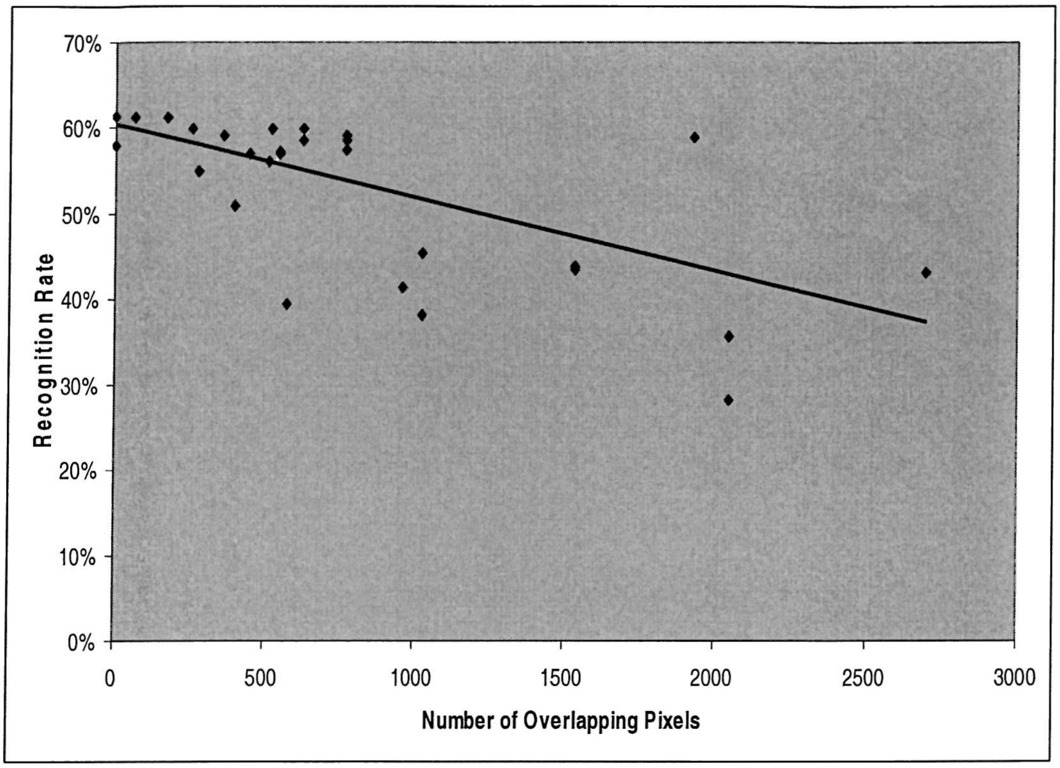


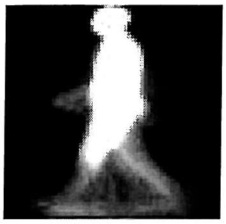
Figure 19 Relationship between Overlapping Pixels and Recognition Rate

By combining the results from additional masks, it is possible to increase the performance further still, however the benefits are negligible. By combining all seven masks, the performance of the system rises to 71.9% which is only a single percentage point higher than that exhibited by combining just a selection of four area masks, as shown in Table 5.

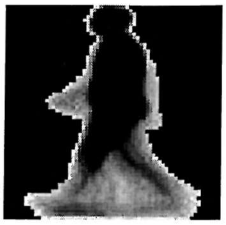
5.3.3 Using only the Temporal Components

Gait differs from traditional biometrics because it contains a temporal as well as a spatial component. The spatial component represents the shape of the person, whilst the temporal component represents the movement of the body over time. Psychological research [8] would seem to suggest that humans can recognise movement patterns merely from the temporal component. This work is backed up by model-based approaches to gait recognition that use the dynamics of movement to recognise subjects [21, 22, 35, 36].

To determine where the recognition capability in our system was derived, we repeated the tests, but this time removing the DC component of gait sequence. This was accomplished by subtracting the mean of each sequence from the data. Thus, the sequence only contains temporal information.



(a) Mean Image for entire Sequence



(b) Sample Temporal Component

Figure 20 Mean and Temporal Component for Sample Gait Sequence

Figure 20(a) illustrates the mean image from a sample gait sequence, and Figure 20 (b) represents a sample temporal component which is obtained by subtracting the mean from the original image. Table 7 shows the recognition rates achieved using only the temporal components of each sequence.

Area Mask	Recognition Rate
Bottom Half (BH)	26.8%
Bottom Triangle (BT)	26.8%
Horizontal Line (HL)	27.6%
Middle Triangle (MT)	24.6%
Top Half (TH)	13.2%
T-shape (TS)	30.6%
Vertical Line (VL)	29.9%

Table 7 Results when DC Component is Removed

The recognition rates range from 13% to over 30%. The performance of the top half mask is significantly worse than that of other masks. This is because the top half of a walking subject consists of very little temporal information, and hence there is insufficient information to discriminate between subjects. By removing the DC component, we are only recognising subjects by the temporal information, rather than the static. Removing the DC component gives some degree of invariance to clothing and ensures we are only recognising subjects by the manner in which they walk, rather than static information such as body shape. Chapter 9 presents preliminary results from the intra-subject variance database and illustrates how using the temporal components alone can negate some of the problems that occur when the subject is wearing different clothing or carrying objects.

By combining the temporal information from multiple masks, it is again possible to increase the performance. By combining the seven masks shown in Table 7 together the recognition results increased to 49.9% This is an interesting result, showing that we can achieve a correct classification result of almost 50%, solely using temporal information cues.

In contrast, we ran experiments where the DC component of each area mask was used as the sole feature. The feature vectors for area masks were calculated on the original images (i.e. with the DC component still present) and only the static DC component of the area mask feature vector was considered as a feature. This gave a single number for each area mask that described the subject in terms of mean area. The recognition rates for individual masks ranged from 0-2% and were very disappointing. However, by combining all DC components from seven masks as above, a recognition rate of 30% was possible. The static information present in a gait sequence is therefore very descriptive as this recognition rate was achieved using just seven numbers.

5.3.4 Using only the Legs

In this section, we present work that uses only the legs to discriminate between subjects. In previous experiments we have shown that area masks which focus exclusively on the top half of the subject perform poorly (see Table 4), and intuitively the legs feature the highest amount of dynamic information in a walking subject. Figure 21 illustrates the area masks chosen to focus exclusively on the legs of the subject.

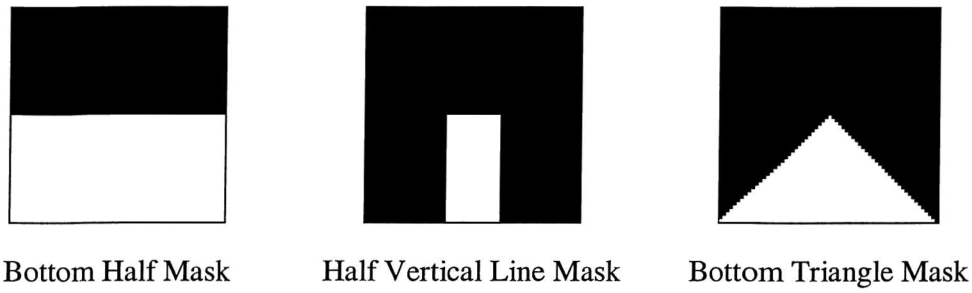


Figure 21 Masks focusing on the legs of subjects

By focusing exclusively on the legs of subjects we gain invariance to a wide variety of conditions. For example, if a subject is carrying some kind of load (such as a backpack) then this is evidenced in large changes in the top half of the silhouette, however the bottom half of the subject does not change (assuming the subject's gait stays the same when carrying a load). Research by Phillips et al [54] has shown that the lower 20% of the silhouette seems to account for 90% of the identification rate in the baseline gait recognition algorithm.

Mask(s) Used	Recognition Rate	
	DC component present	DC component Removed
Bottom Half Mask	35.6%	26.8%
Bottom Triangle Mask	38.2%	26.8%
Half Vertical Line Mask	39.8%	25.2%
All 3 combined	53%	38.9%

Table 8 Using only the Legs

The recognition rates for individual masks are shown in Table 8. As can be seen, the performance level from each individual mask is similar. The recognition rate drops to approximately 25% when the DC component (static information) is removed from the original input silhouettes. By combining together all three masks shown in Table 8 (with the DC component present) the recognition rate improved to 53%. Using only the temporal components from each mask and combining them together results in a recognition rate of 38.9%. In contrast, using the DC components alone (with the three area masks illustrated in Figure 21) gives a recognition rate of 37.8%, using just three numbers to describe each gait sequence.

If we compare the results using just the legs to those using the silhouettes we see that a large proportion of the performance of the system comes from the extra information that is available in the torso region of the subject. However, the results using the DC components alone are interesting, with the DC components of the legs giving better performance (37.8% compared to 30%) than the DC components when the whole silhouette is considered.

5.4 Using Moment Based Descriptors

5.4.1 Person Identification using a Single Moment

As a baseline measure of moment based descriptor performance, we used a single moment invariant (without using a masking circle) to describe each frame. The recognition results quoted are upon the SOTON database with 114 subjects and eight samples each. Leave one out cross validation was used, together with a k nearest neighbour classifier. Table 9 shows the performance using a selection of Hu moment invariants. The feature vector from each moment was combined by simply concatenating the feature vectors of each individual moment invariant to form one large feature descriptor.

Moment Invariant	Recognition Rate
Hu1	44.7%
Hu2	35.9%
Hu8	22.7%
All 3 combined	51.4%

Table 9 Performance of Moment Invariants for Person Discrimination

Performance levels using a single moment invariant are similar to those of using a single area mask shown in Table 4. By combining the feature vectors, the performance level was raised, indicating independent information from each of the moment descriptors.

5.4.2 Person Identification using a Family of Shape Descriptors

By using a family of shape descriptors, we can describe the shape with more accuracy and thus increase the performance of the description and therefore the recognition rate. Table 10 demonstrates how we improve performance using a family of descriptors.

Values of α	Recognition Rate			
	Hu1	**Hu2**	**Hu8**	**Combined**
0 (no masking circle)	44.7%	35.9%	22.7%	51.4%
0 to 0.5 in steps of 0.05	63.3%	47.6%	29.1%	63.3%
0 to 1 in steps of 0.1	68.3%	44.5%	28.5%	56.7%
0 to 1.5 in steps of 0.25	65.9%	34.3%	28.8%	42.1%
0 to 2 in steps of 0.1	64.4%	39.4%	31.6%	44.7%

Table 10 Using a Family of Shape Descriptors for Person Discrimination

As can be seen from Table 10 the performance using a family of shape descriptors rises rapidly when low values of α are used but the performance falls away when large values of α are combined. This suggests that using large masking circles and thus describing information at the periphery of the silhouette can detract from recognition performance.

The Hu1 moment performs consistently better than other moments at discriminating between subjects. Both the Hu2 and Hu8 moments perform poorly, achieving low recognition rates despite of the increased information present from using masking circles. The major difference between the Hu2 and Hu8 moments is the inclusion of the M_{11} term. We conclude from this that the M_{11} term is not very descriptive for discriminating between subjects because the moments that contain this term perform significantly worse than those that do not.

5.5 Discussion

This chapter has presented results using two new techniques for automatic gait recognition, area masks and moment based descriptors. Area masks recognise a subject by a dynamic signature that describes how the area changes in a particular portion of the image. In this chapter, we have shown that recognition rates from individual masks vary from 28% to 48%. By utilising gaits' many degrees of freedom by combining the results from different masks the performance rises sharply. By combining any pair of masks, the performance ranges from 45% to 65%. We hypothesise that the amount of overlap between two masks is inversely related to the recognition performance. If more masks are considered the performance rises further still. If all seven masks are combined then the recognition rate rises once again to 71.9%

The temporal component of gait is unique in the field of biometrics. By removing the static (DC) component we have shown that one can achieve a recognition rate of 50% by just considering the temporal components of gait alone, rather than static cues such as body shape. In contrast, if the temporal component of gait is not considered then the recognition rate is only 30%.

Moment based descriptors give a geometric description for each silhouette in the sequence. By using a single moment descriptor for each frame in the image recognition rates range from 22.7% to 44.7%. A family of shape descriptors provides additional information about the gait sequence, and increases performance. Masking circles are used to define portions of the image to describe, and using a family of these circles provides better description. By using a family of descriptors, the performance increase was substantial rising to 68.3% using the Hu1 invariant alone.

In comparison with the Hu1 moment, experimental results have shown that the Hu2 and Hu8 moments are poor discriminators between subjects. When a family of masking circles was used the highest recognition performance for the Hu2/Hu8 moments was still less than 50%. The Hu2 and Hu8 moments both contain the M_{11} term. We conclude that this term is a poor discriminator between subjects.

In conclusion, in this chapter we have presented two new techniques for automatic gait recognition. By using area masks, and combining the information together from multiple masks, a recognition rate of over 70% was achieved on a database consisting of 114 subjects. Results using moment based descriptors show that by using a single moment descriptor a recognition rate of over 40% can be achieved. Using a family of shape descriptors greatly improves the recognition performance, and a recognition rate of 68.3% was achieved on a database of 114 subjects.

In the following chapter, we look at the task of gender discrimination using both area masks and moment based descriptors.

Chapter 6 Gender Discrimination

6.1 Introduction

Gender discrimination could be a practical first stage in automatic gait classification if the number of subjects is large, as it would reduce the number of subjects needing to be searched. Gender discrimination is also an interesting area for research, as previously psychological studies have indicated that discriminatory cues are best discerned from fronto-normal views rather than the sagittal views used here [10, 11]. The database used in these experiments is a subset of the SOTON database used in previous tests. The SOTON database consists of 114 subjects, of which 24 are female. To avoid bias in the database, 24 people of each sex were used (all the females and a random selection of men). A single sequence of each subject was used to create a database consisting of 48 individual sequences. By using a single sequence from each subject, we are ensuring that we are discriminating by gender, rather than discriminating between subjects. We shall call this database the GEN-SOTON database. Further details about the GEN-SOTON database, and the subjects used in it can be found in Appendix A. A k nearest neighbour algorithm was employed, and leave one out cross validation was used in all of the following experiments.

6.2 Area Masks

Table 11 shows the recognition rates for individual masks, using the GEN-SOTON database in the manner described above. As can be seen the recognition rates range from 41.7% to 77.1%. The chance recognition rate on the database is 50%, and most masks perform better than that, with the exception of the top half mask, which is providing no gender discrimination information.

Mask Used	Recognition Rate
Bottom Half (BH)	64.6%
Bottom Triangle (BT)	56.3%
Horizontal Line (HL)	66.7%
Middle Triangle (MT)	77.1%
Top Half (TH)	41.7%
T-shape (TS)	54.2%
Vertical Line (VL)	62.5%

Table 11 Gender Discrimination Rates for Individual Masks

By combining the results from the individual masks, the recognition rate was 60.4%. Previous medical work has actually indicated that male and female walkers differ in terms of lateral body sway, with males tending to swing their shoulders from side to side more than their hips, and females tending to swing their hips more than their shoulders [10, 11]. This sort of information is better discerned from a fronto-normal or an overhead view, which explains the poor performance here when masks are combined. There are differences in the static information between male and female, but the resolution of the silhouettes (64 x 64) may be too poor to determine this information. In addition, by normalising for height we have removed a known gender discriminator. Further research is necessary to determine the gender discrimination characteristics of each mask.

6.3 Moment Based Descriptors

Moment based descriptors describe a gait sequence in terms of geometrically invariant properties. Are these able to better discern the sex of a walker? The GEN-SOTON database was used, and the alpha range chosen was between zero and one inclusive, in steps of 0.1.

Moment Invariant	Recognition Rate
Hu1	52.1%
Hu2	68.8%
Hu8	66.7%
All 3 combined	64.6%

Table 12 Gender Discrimination using Moment Based Descriptors

Table 12 shows the gender discrimination rates using moment-based descriptors. The Hu1 moment performs poorly with a recognition rate only slightly better than that of chance (50%). The Hu2 and Hu8 moments performs best, which is an interesting result as in the case of person identification, both moments perform poorly (see Table 9).

Due to the performance differences with the moments, we re-examined the effects of using the masking circle. Table 13 shows the results of using various sizes of masking circles.

Values of α	Recognition Rate			
	Hu1	Hu2	Hu8	Combined
0 to 0.5 in steps of 0.05	64.6%	62.5%	64.6%	56.3%
0 to 1 in steps of 0.1	52.1%	68.8%	66.7%	64.6%
0 to 1.5 in steps of 0.25	64.6%	66.7%	72.9%	64.6%
0 to 2 in steps of 0.1	60.4%	75.0%	64.6%	70.8%

Table 13 Using a Family of Shape Descriptors for Gender Discrimination

By altering the alpha range to include high values of alpha, the recognition rate increased to 75% when the Hu2 moment was used. Further research is necessary to determine the properties of the Hu2 moment that lead it to yield increased performance.

6.4 Discussion

This chapter has presented results on the task of gender discrimination from both area masks and moment based descriptors. Area masks achieve gender discrimination rates between 41.7% and 77.1%. The highest recognition rate came from the mid triangle mask illustrated in Figure 5 (g). This mask isolates the hips and top half of the torso and illustrates that there is significant statistical difference in the dynamics of area change in this part of the body during a gait cycle for different genders.

The performance using moment based descriptors also yields some interesting results. In the case of discriminating between subjects, the Hu1 moment consistently yielded the best results. When discriminating between genders the Hu2 and Hu8 moments performed significantly better than the Hu1 moment, with a maximum recognition rate of 75%. The major difference between the Hu1 moment and the other two moments is the inclusion of the M_{11} term in the calculations (see equation (4-6)). This implies that the M_{11} term provides additional gender discrimination information.

The next chapter presents work on whether gait is a symmetric pattern of movement. We have considered gait a symmetric pattern of movement and thus not considered heel strike information or direction of travel. If gait is asymmetric, then by considering the heel strike information we should be able to increase the recognition rates.

Chapter 7 Is Gait Symmetric?

7.1 Introduction

Gait symmetry has been defined as the perfect agreement between actions of the lower limbs [55, 56]. Gabbard [57] has suggested using the term when no statistical differences are noted on parameters measured bilaterally. Historically, the psychologists' view is that gait is a symmetrical pattern of motion [58, 59]. So far, we have assumed that this is true, and do not take into account the foot on which the subject starts, or the direction of travel. Does taking this information into account result in a significant difference in recognition rate? The SOTON database has labelled information including heel strikes (i.e. the foot the subject starts a gait cycle with) and direction of walk. By using this information, we were able to undertake experiments where the direction of walk and the starting foot for the gait cycle were taken into account.

7.2 Using Area Masks

By taking into account the direction of walk and/or the starting foot of the gait cycle are we able to improve performance? Table 14 shows the recognition results from the SOTON database when the starting leg and direction is considered. Seven area masks (BH, BT, HL, MT, TH, TS,VL) were used and combined and leave one out cross validation was used to produce the results using a k nearest neighbour algorithm. As can be seen, by considering the direction of walk and starting foot and constraining the database accordingly, it is possible to improve the recognition results dramatically.

Direction of Walk	Starting Foot	Recognition Rate
Left	Left	84.3%
Left	Right	86.2%
Left	Both left and right	89.0%
Right	Left	84.9%
Right	Right	81.3%
Right	Both left and right	86.5%
Left and Right	Left	85.0%
Left and Right	Right	84.4%

Table 14 Recognition Rates when Considering Direction and Start Leg using Area Masks

In all circumstances the recognition rate is over 80%, which compares to a best performance of just 71.9% when direction of walk and starting foot are not considered. Table 15 shows the recognition rates for single masks when the direction of travel and the starting heel strike leg are the same. The results are taken from when the subject is walking left to right and the left heel starts the gait cycle.

Mask Used	Recognition Rate
Bottom Half (BH)	54.9%
Bottom Triangle (BT)	52.6%
Horizontal Line (HL)	60.1%
Middle Triangle (MT)	65.2%
Top Half (TH)	44.9%
T-shape (TS)	62.9%
Vertical Line (VL)	62.4%

Table 15 Recognition Rates when Direction and Start Leg are the same

As can be seen from Table 15 the recognition rates are considerably higher than those when the whole database is considered. Figure 22 illustrates the performance difference between using the whole database, to using a constrained portion of the database where the direction of travel and starting leg are the same. On average the performance advantage (using a single area mask) of using the extra information of direction of travel and starting leg is approximately 50%

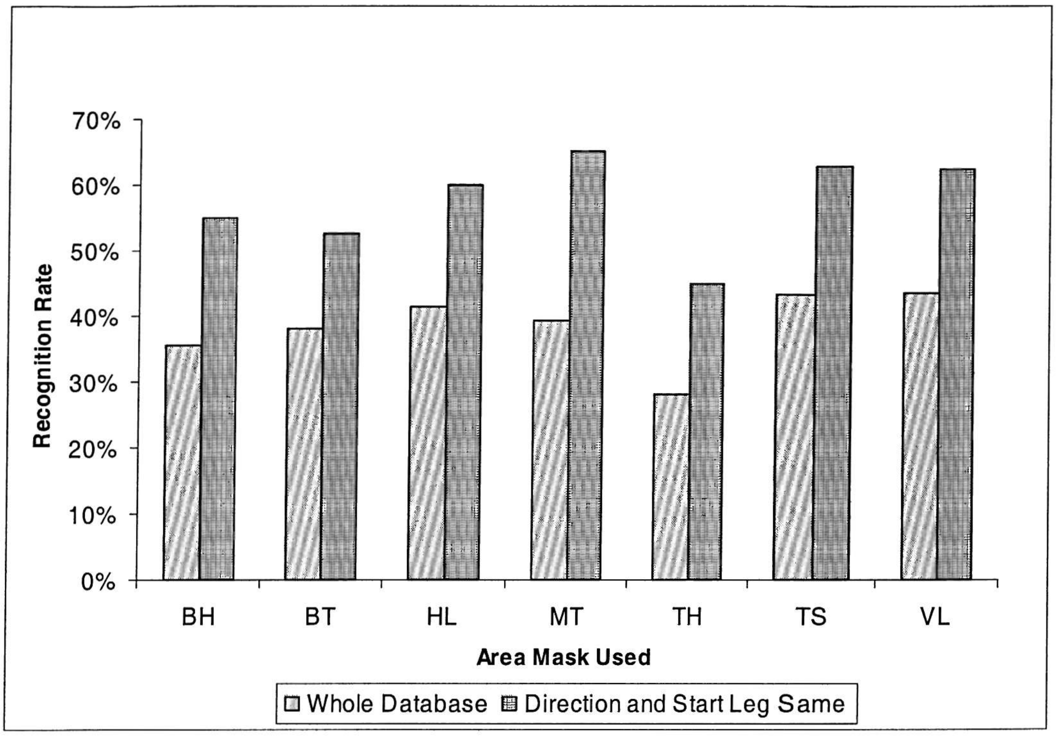


Figure 22 The Performance Difference when Starting Leg and Direction are Considered

In section 5.2 we explained that the poor performance demonstrated when using higher values of k could be because of poor clustering in feature space due to disregarding heel strike and direction of travel information. In the following experiment we used 114 subjects but only considered sequences where the subject began the gait cycle on the left heel. Various values of k were used for classification.

Area Mask	Recognition Rate		
	k=1	k=3	k=5
Bottom Half Mask	52.1%	36.0%	24.9%
Top Half Mask	45.2%	30.7%	21.6%
Vertical Line Mask	61.8%	46.3%	36.2%
Seven masks combined	85.0%	74.3%	63.8%

Table 16 Effects of *k* in Classification when Heel Strikes are Considered

As can be seen from Table 16 the performance decrease with larger values of k is less marked than when heel strike information was not considered. If we compare the results to those in Table 3, we can see that the performance level is substantially higher for all values of k.

When we chose which classifier to use, we rejected using canonical analysis because of the poor performance and computational complexity. Is the performance of the classifier improved if the starting leg and heel strike information is considered? In the following table, canonical analysis was used to provide the recognition rates on a variety of area masks. Only sequences with left heel strikes and walking left to right were considered. Canonical analysis was used in the manner described previously in section 5.2.

Mask(s) Used	Recognition Rate
Bottom Half (BH)	32.8%
Horizontal Line (HL)	31.4%
Vertical Line (VL)	39.6%
Seven masks (BH, BT, HL, MT, TH, TS, VL)	76.3%

Table 17 Canonical Analysis Performance when Heel Strike and Direction of Travel Information is Considered

As can be seen from Table 17, the recognition performance when using canonical analysis is greatly increased when heel strike information is considered. This suggests that the clustering of data points is greatly improved in canonical space when heel strike information and direction of walk information is used to constrain the database.

7.3 Using Moment Based Descriptors

In the previous section we have seen how by considering the walk direction and starting leg, we were able to markedly increase the performance using area masks for gait recognition. In this section, we perform the same experiments, but using moment based descriptors. The range of masking circles in this experiment was between zero and one inclusive in steps of 0.1.

Direction of Walk	Starting Foot	Recognition Rate using Hu1 Moments
Left	Left	83.2%
Left	Right	80.6%
Right	Left	81.9%
Right	Right	78.7%
Left	Both left and right	86.9%
Right	Both left and right	83.2%
Left and Right	Left	82.9%
Left and Right	Right	81.5%

Table 18 Recognition Rates when Considering Direction and Start Leg using Moment Based Descriptors

As can be seen from Table 18 the performance gain when considering the starting leg is substantial. This suggests, once again, that there is an asymmetrical component of gait and by taking this into consideration performance can be improved.

For area masks we have illustrated that when the direction of travel and heel strikes are considered we can improve the clustering in feature space. Table 19 illustrates the effects of k on recognition performance when we constrain the database to specific heel strike and direction of travel.

Value of k	Recognition Rate using Hu1 Moments
1	82.9%
3	71.2%
5	56.8%

Table 19 Effects of *k* on Recognition Performance when Heel Strikes are Considered

As can be seen from Table 19, the performance degradation when using high values of k in the classification algorithm is markedly less than that exhibited when heel strike information is not considered. This suggests that the clustering in feature space is much improved.

7.4 Discussion

In a recent study, Sadeghi et al. [59] examined the assumption made in the literature that gait is symmetric. Asymmetric gait is frequently considered to indicate gait pathology. Studies by Singh [60] and Du Chatinier et al [61] have supported the notion of asymmetry in the normal population. Sadeghi et al. [59] suggest that limb dominance affects the asymmetry of gait. In this chapter we have performed experiments whose results show that gait does indeed have an asymmetric component. The improvement brought about by considering this component is substantial, suggesting that the asymmetric component of gait is a good discriminator between subjects.

As has been shown, the performance gain when considering the start leg and direction of travel is high. This suggests that there is an asymmetrical component to gait that has high inter-class variance and thus increases recognition performance. The performance increase using area masks when start leg and direction of travel is considered is high, providing an average of a 50% performance gain for individual masks. Using moment based descriptors the performance gain is also high. The recognition rates increase from 68% to an average of approximately 80% when the direction of walk and starting leg of the gait cycle is considered.

In this chapter, we have shown that by considering the starting heel strike for the gait cycle the performance increases dramatically. Even by simply considering the direction of walk, the recognition performance can be increased considerably. This work confirms previous studies that suggest gait has asymmetric components due to limb dominance. Further to this it suggests that these asymmetric components provide good discrimination between subjects.

We have also shown that the clustering of subjects in feature space is noticeably improved if direction of travel and heel strike information is considered. This is evidenced in the improved performance with canonical analysis and the better performance using higher values of k when using the k nearest neighbour classification algorithm.

In the following chapter, we look at performance of area masks and moment based descriptors under adverse conditions.

Chapter 8 Performance Analysis

8.1 Introduction

In previous chapters, we have reported the results where we have used the SOTON database. The part of the SOTON database we have used is derived in the laboratory and thus the results obtained are not indicative of performance in real world conditions. This chapter looks at the generalisation properties of both area masks and moment based descriptors.

In "real-world" conditions, extracting a noise-free silhouette is a difficult problem. To simulate the effects of noise, we have added noise in varying degrees to the database and tested how each technique responds. Although the noise added is not realistic, it can give us some idea of how the technique will cope with imperfect silhouettes.

One of the chief advantages of gait recognition is that it allows recognition from a distance. One of the problems with this is that it may result in only low resolution silhouettes of a subject being available. To see how both techniques would cope with low-resolution silhouettes, we rescaled the silhouettes down in varying degrees from their original size of 64 x 64. In addition, we tested the rotational invariance properties of moment based descriptors and also the sensitivity to image centring of the area masks algorithm.

8.2 Area Masks

8.2.1 Noise Analysis

The extracted silhouettes used in the performance analysis are derived under laboratory conditions and are thus not indicative of performance under "real-world" conditions. To simulate the effects of noise we added "salt and pepper" noise in varying degrees to the database. This noise is not a realistic simulation of the noise that would be encountered in practise (and could easily be reduced), but it does give an idea of the generalisation capabilities of area masks.

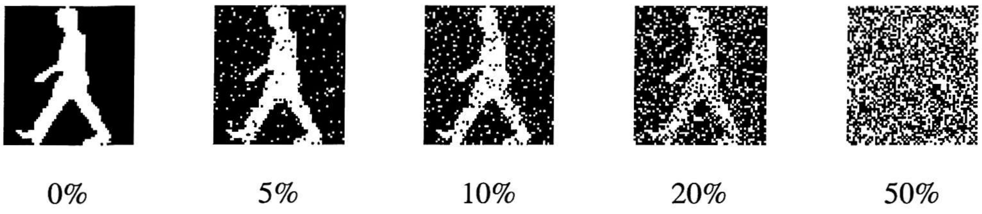


Figure 23 Sample Noisy Images

The noise level reflects the percentage of pixels that are changed out of the total number of pixels. Therefore, a noise level of 50% means that half the pixels in the image are flipped from their previous binary value. As can be seen from Figure 23 once the noise level reaches high values it is impossible to discern any human silhouette in the image whatsoever.

The entire SOTON database was used for the test. A *k*-nearest neighbour was used to provide the results and leave one out cross validation was used. Seven area masks were combined together and one gait cycle was used. Thirty samples were taken from each feature vector using a cubic spline curve. The seven area masks combined were Bottom Half, Bottom Triangle, Horizontal Line, Middle Triangle, Top Half, T-shape, and Vertical Line.

Amount of Noise Added to System	Recognition Rate
0%	71.9%
1%	71.5%
2%	71.4%
5%	69.3%
10%	66.0%
15%	57.6%
20%	48.1%
25%	37.1%
30%	25.2%
35%	13.1%
40%	6.0%
45%	2.1%
50%	0.8%

Table 20 Effects when Noise is added to the System using Area Masks

As can be seen from Table 20, the recognition rate drops dramatically for high noise values. In fact, when the noise level reaches 50% the recognition rate is approximately equal to that given by chance. At low levels of noise, up to approximately 10%, the recognition rate is respectable at approaching 70%.

8.2.2 Errors in Windowing the Subject

Area masks are sensitive to being centred on the centre of mass of the image. What happens when the area mask is not centred on the image? To simulate the effects of this we created versions of the vertical line mask, and these were misaligned to various degrees. Figure 24 shows some sample misaligned vertical line masks. The numeric values in the table refer to how far the mask has been shifted from a central position.

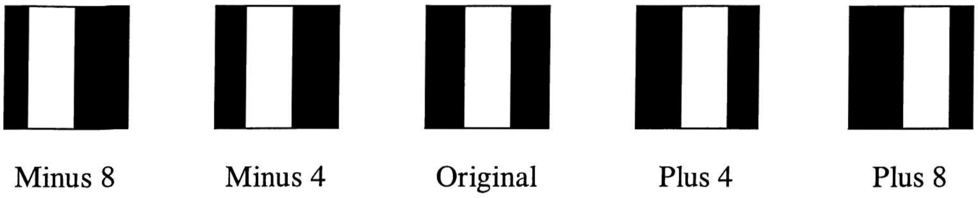


Figure 24 Misaligned Vertical Line Masks

Using the misaligned vertical line masks, we ran experiments on the entire SOTON database to examine the performance issues associated with poorly centred images.

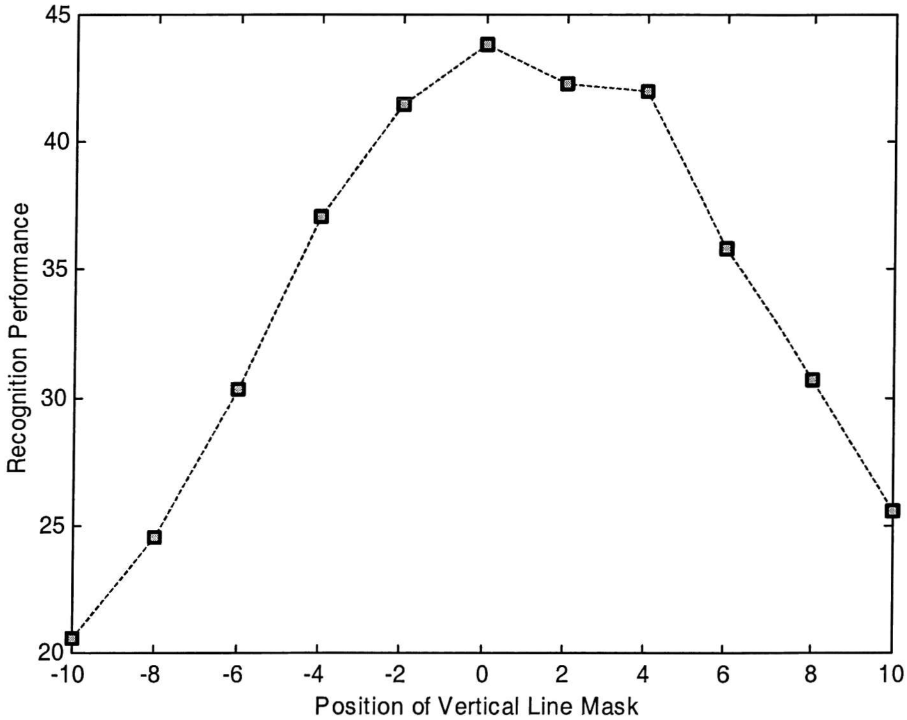


Figure 25 Recognition Performance for Shifted Images

Figure 25 illustrates the recognition performance when the collection of shifted vertical line masks was used. As can be clearly seen from the graph, the maximum performance occurs when the mask is centred and hence invariant to left-right flipping. As the absolute position of the mask increases from the centre, the recognition rate drops significantly.

8.2.3 Scaling the Silhouettes

One of the primary advantages of gait as a biometric is that it allows recognition from a great distance. To simulate viewing video from a distance, the silhouettes were rescaled downwards from their original size of 64 x 64. Figure 26 illustrates several sizes of silhouettes used in the experiments. As can be seen, as the resolution reduces, the silhouettes become increasingly blocky and less fine detail can be determined from the images.

Figure 26 Rescaled Silhouettes

We ran experiments upon the SOTON database as before, using 114 subjects with 8 samples of each subject. The quoted recognition rate uses seven area masks combined, using a *k* nearest neighbour algorithm in feature space.

Size of Input Silhouettes	Recognition Rate
64x64	71.9%
56x56	70.9%
48x48	72.2%
40x40	69.3%
32x32	69.2%
24x24	68.5%
16x16	60.3%
8x8	47.3%

Table 21 Effects of Scaling using Area Masks

Seven area masks were used, as illustrated in Figure 5 (excluding the 'Full Mask'). As can be seen from Table 21, the performance does not degenerate significantly until very small silhouette sizes are used such as 16x16 or 8x8. Figure 27 shows output from the vertical line mask from a single subject at different resolutions. The feature vector for each scale is normalised by dividing it by the area of the input image for the purposes of visualisation. By doing this, we make area masks invariant to scale.

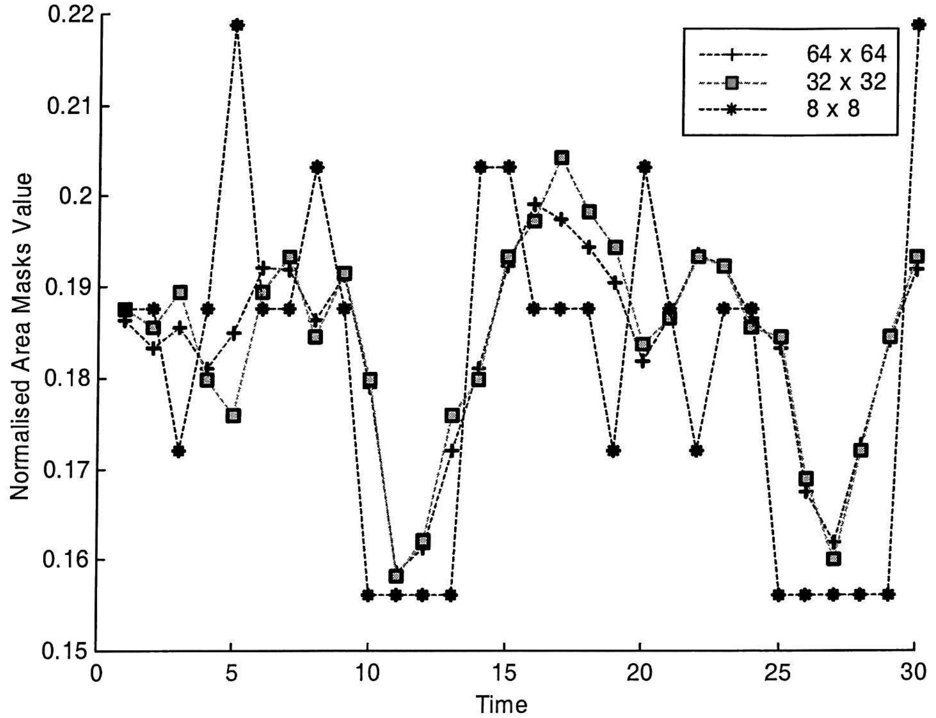


Figure 27 Effects of Scaling on Vertical Line Mask

As can be seen from the graph, even at different scales the overall "shape" of feature vector remains constant, with the 64 x 64 output being almost indistinguishable from the 32 x 32 output. When the resolution of the images drops to 8 x 8 the overall "shape" is still present, but the feature vector is very noisy, explaining the poor performance.

8.2.4 Effects of Rotation

Rotated silhouettes may occur due to poor camera calibration or from gathering information from real-world sources (e.g. CCTV cameras). How do area masks deal with rotation? To compare the moment descriptors and area masks, we used a set of silhouettes rotated at various angles. Figure 28 illustrates a sample selection of rotated silhouettes.

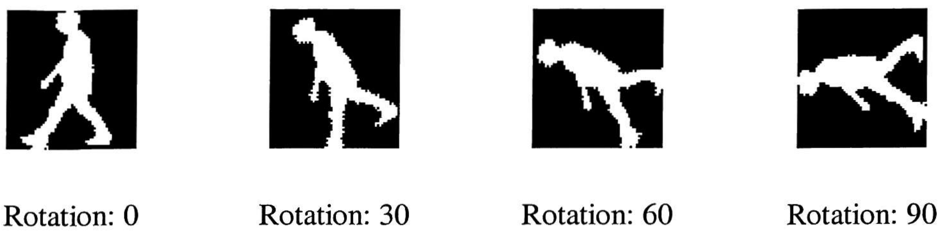


Figure 28 Rotated Silhouettes

Figure 29 illustrates the changes the feature vector undergoes for various angles of rotation. The illustrated feature vector is the vertical line mask, shown in Figure 5 (b).

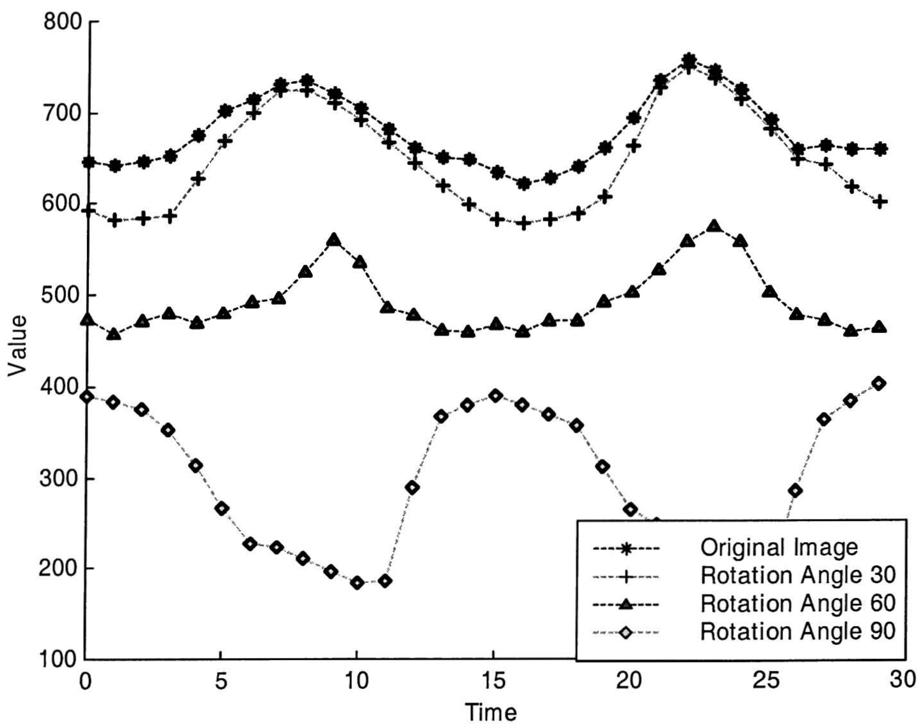


Figure 29 Effects of Rotation on Area Masks Feature vector

As is clear from Figure 29, area masks are not invariant to rotation as the feature vector changes substantially when the silhouettes are rotated. Moment based descriptors have an advantage over area masks in this respect in that they are invariant to rotation (i.e. the feature vector describing the same subject at different angles of rotation is almost identical). The effects of rotation using moment based descriptors are examined more closely in section 8.3.3.

8.3 *Moment Based Descriptors*

8.3.1 Noise Analysis

To test the generalisation capabilities of the moment based descriptor approach to gait recognition we added noise at various levels to the original input silhouettes (in the same manner described in the section 8.2.1). Once again, the noise added is not realistic, but it does give some idea of the generalisation capabilities of the moment based descriptor technique. In Table 22, the Hu1 moments were used with an alpha range of between 0 and 1 inclusive in steps of 0.1.

Amount of Noise Added to System	Recognition Rate
0%	68.3%
1%	68.0%
2%	64.0%
5%	60.2%
10%	47.6%
15%	37.9%
20%	24.6%
25%	18.8%
30%	11.1%

Table 22 Effects when Noise is added to the System using Moment Descriptors

As can be seen from Table 22, moment based descriptors are more sensitive to noise than area masks. The performance level remains high at over 60% for levels of noise up to approximately 5%. When the noise level is increased further the performance of moment based descriptors drops rapidly.

8.3.2 Scaling the Silhouettes

By using Hu invariants, our moment based technique should be invariant to scaling. However, due to quantisation errors, this is only approximately true in the case of digital images. We ran experiments using the scaled down silhouettes shown in

Figure 26 using moment based descriptors. Hu1 moments were used, with an alpha range of between zero and one inclusive in steps of 0.1.

Size of Input Silhouettes	Recognition Rate
64x64	68.3%
56x56	69.4%
48x48	67.7%
40x40	67.9%
32x32	64.4%
24x24	62.0%
16x16	52.1%
8x8	27.6%

Table 23 Effects of Scaling using Moment Based Descriptors

Table 23 illustrates the effects of on recognition rates when the images are scaled. As can be seen the performance loss is not significant until the image is resized to small sizes of 16x16 or 8x8. Note that the scale does not need to be altered for visualisation purposes since the moments used are invariant to scale. If we compare the performance between moment based descriptors and area masks, we see that area masks offer better performance when using low resolution silhouettes. However, moment based descriptors have the advantage of being invariant to scale.

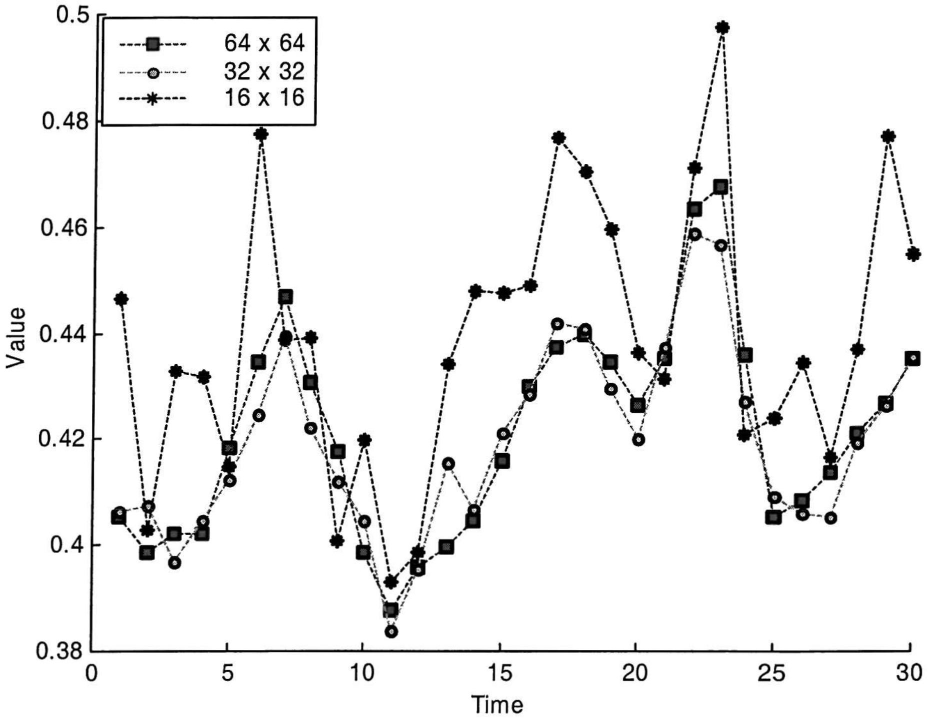


Figure 30 Hu1 Moments for a Single Subject at varying Resolutions

Figure 30 illustrates graphically the affects of scale on the moment values. In the graph, the Hu1 moment is illustrated, with alpha having a value of zero. As can be seen the moment values for 64 x 64 and 32 x 32 silhouettes are almost the same. However, when the silhouette size is reduced to 16 x 16 it is difficult to compare the two plots (64x64 and 16x16 in Figure 32).

8.3.3 Invariance to Rotation

Hu invariant moments are invariant to rotation, scaling and translation. For digital images, this invariance is only approximate. To evaluate the performance of moment based descriptors, we rotated the images through various angles and ran the tests as before. Figure 28 illustrates a sample set of rotated silhouettes.

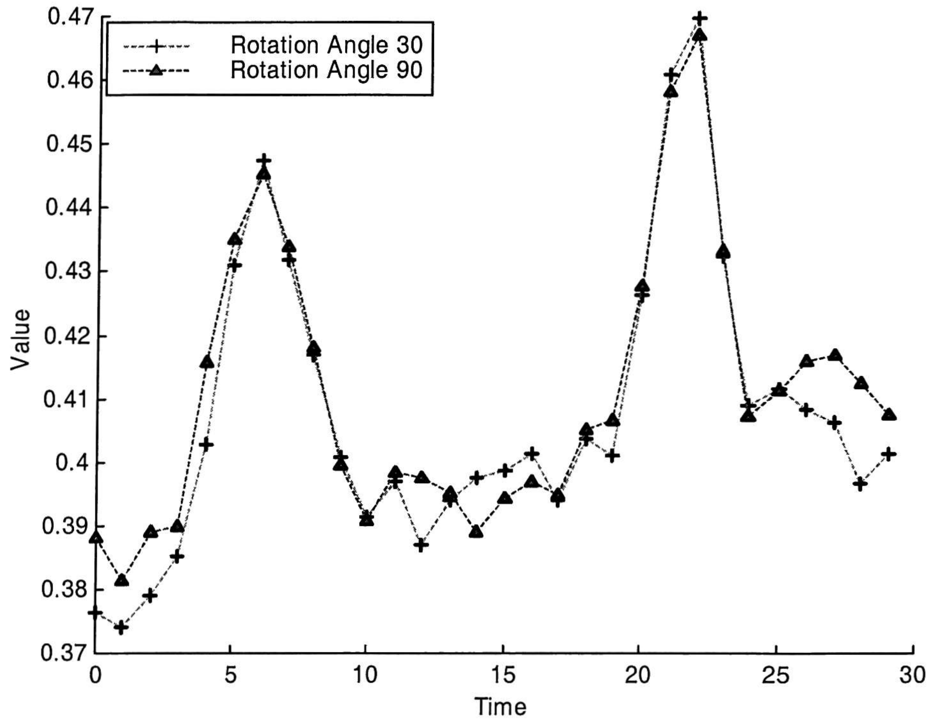


Figure 31 Differences due to Digitisation errors

Figure 31 illustrates the differences between the moment descriptors when the silhouette is rotated at two different angles. As can be seen the general shape of description is very similar, but small differences in value occur due to digitisation errors. The small differences in recognition performance indicate that the technique is "almost" invariant to rotation of the silhouette, but due to digitisation small errors will always exist. Area masks are unable to cope with rotated silhouettes whereas moment based descriptors provide (almost) invariance to rotation.

8.4 Discussion

In this chapter, we have examined the performance characteristics of area masks and moment based descriptors. The most basic performance characteristic is how a given technique responds to noise being added to the original images. The noise added is not realistic, but it does give some indication of the generalisation capabilities of each technique. Figure 32 illustrates the performance of both techniques when varying degrees of noise are added to the original images.

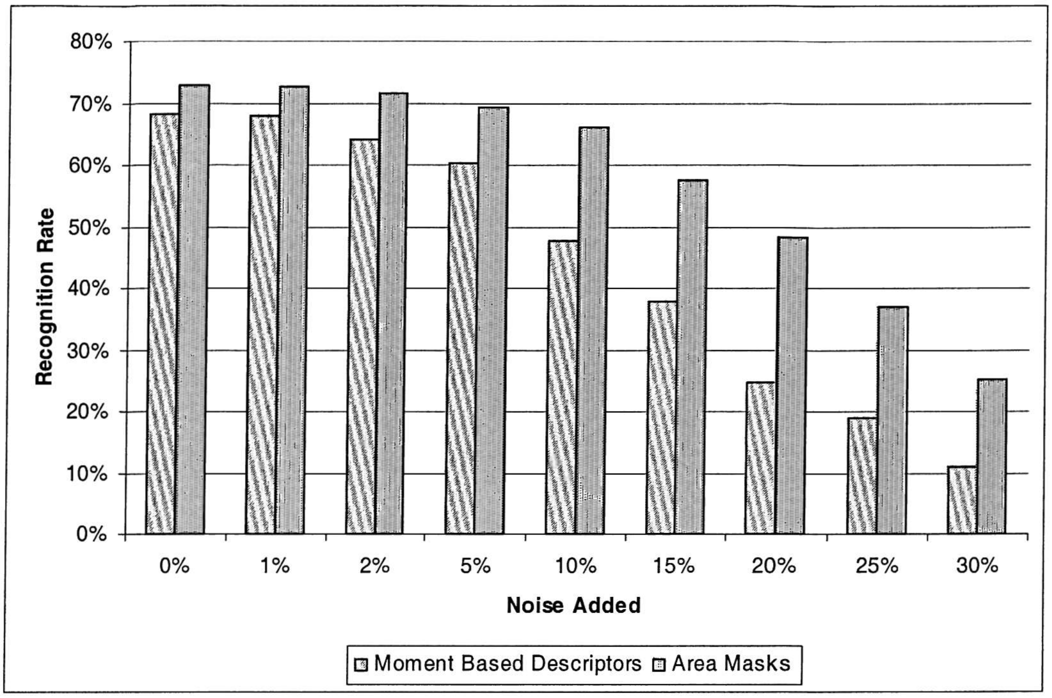


Figure 32 Performance of both Techniques when Noise is Added

As expected, the performance of both techniques degenerates as increasing levels noise are added to the images. In all instances, area masks perform better than moment based descriptors. The noise results indicate just how important the silhouette extraction process in automatic gait recognition is in achieving high recognition results.

The area mask technique relies on subjects being centred inside a windowed image. This is to ensure that the masks are placed correctly over the human silhouette shape. To simulate the affects of subjects being misplaced inside a windowed image, we created several offset masks and analysed the results on the SOTON database. The results indicate that having the subject correctly centred in the image is vital, as shifting the image by even a small amount can drop the recognition rate dramatically. This again illustrates just how important the silhouette extraction procedure is in producing high recognition results when gait is used as a biometric.

One of automatic gait recognition's chief advantages is that it allows recognition from a distance. Recognition from a distance involves additional challenges because the resolution of the human silhouette is lower. To simulate the affects of lower resolution silhouettes we rescaled the silhouettes in the SOTON database as illustrated in Figure 26.

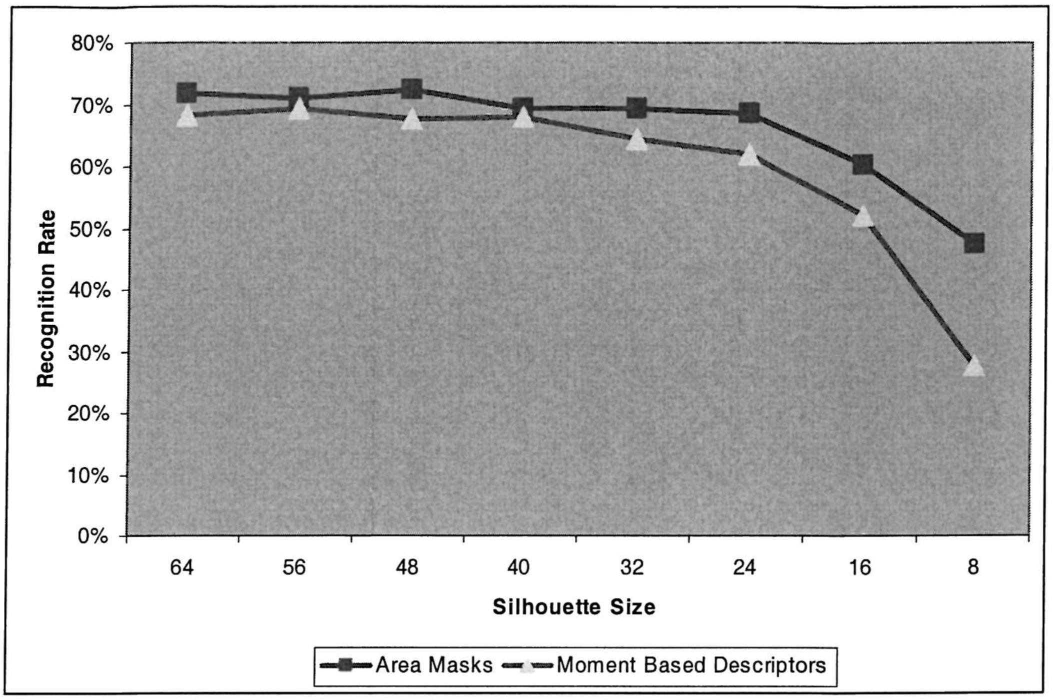


Figure 33 Performance of both Techniques using Rescaled Images

Figure 33 illustrates the performance of both techniques using rescaled silhouettes. The performance of both techniques using low resolution silhouettes is encouraging. Even when the size of the silhouette is one quarter of its original size (32 x 32) both techniques perform within a few percent of their original performance. As the size of the silhouettes drops the performance does decrease, but even the recognition rates using 8 x 8 silhouettes are significantly better than chance in both cases, which provides a powerful illustration of the vast amount of information in a gait sequence.

Rotated silhouettes is a situation that may occur in real-world situations. Moment based descriptors are invariant to silhouette rotation and hence have a large advantage over area masks which are not invariant to rotation.

In Chapter 10, we sum up what has been learnt from this research and illustrate new ideas for further research into automatic gait recognition.

Chapter 9 Intra-Subject Variance

9.1 Introduction

This short chapter presents preliminary results from a new database that aims to address the area of intra-subject variance. The new database is designed for investigation of gaits covariates, such as the effect of different footwear, clothing and load carrying. Subjects are filmed as previously seen in the SOTON database (under laboratory conditions walking normal to the camera's plane of view).

The new database aims to discover how much a subject's gait varies due to different foot-wear, carrying loads and wearing different clothes. We examine the consequences of just using the temporal component of gait and show how this can deliver obvious advantages over using the spatial component. Further research is necessary to examine the consequences of intra-subject variance on recognition in application scenarios.

9.2 Results

9.2.1 Clothing

Clothing is one of the covariates of foremost interest in typical application scenarios. In this section, we look at the resultant feature vectors when a subject is wearing different types of coat, compared with the original silhouettes.

(a) Original (b) Raincoat (c) Trench Coat

Figure 34 Sample Silhouettes from the Intra-Subject variance database

As can be seen from Figure 34, the change to the silhouette is large when the subject is wearing different types of coat. The raincoat dramatically changes the top half of the silhouette, whilst the legs remain relatively unchanged. The trench coat changes the whole of the silhouette and makes the relative motion of the legs difficult to discern. Figure 35 illustrates the results from plotting a single gait cycle using the bottom half mask for each of different types of clothing (original, raincoat, trench coat).

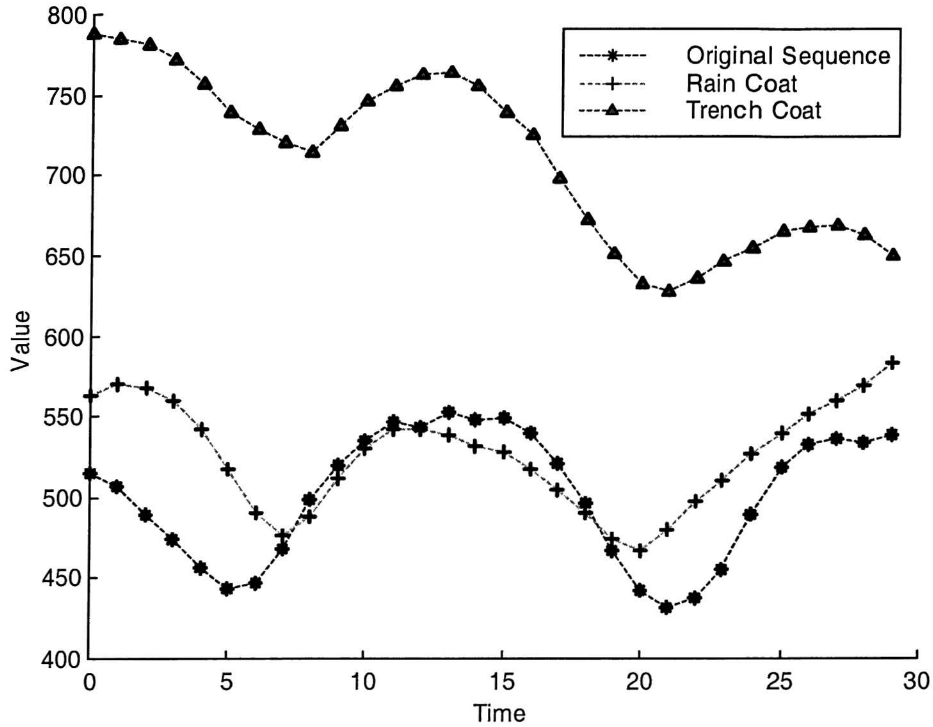


Figure 35 Clothing affects on the Feature Vector from the Bottom Half Mask

As expected, the feature vectors from the bottom half mask are very similar for the original sequence and the raincoat sequence. The trench coat sequence clearly differs in magnitude and the first half of the gait cycle is of larger magnitude than the remainder. This may be to do with the relative motion of the trench coat with respect to human gait. Further research is necessary to determine whether this is so. Previous experiments have indicated that we can achieve a recognition rate of approximately 50% using solely the temporal components of gait. We repeated the experiments, but this time removing the DC component of the original silhouettes. Figure 36 illustrates the results when the DC component of the original silhouettes is removed. As can be seen from the graph, the feature vectors become much more similar when only the dynamic components of gait are considered.

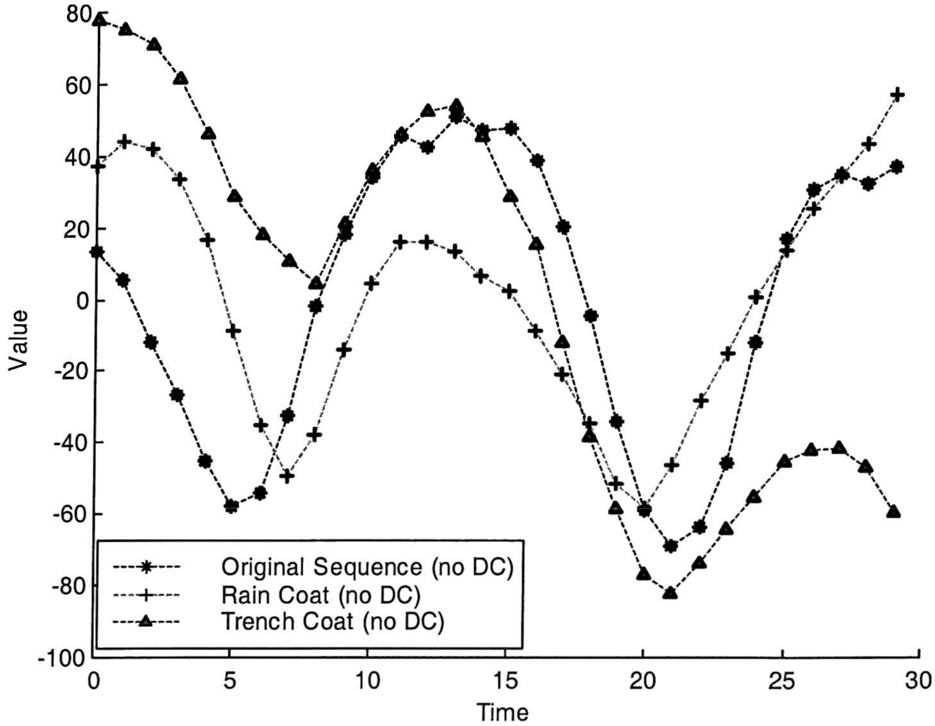


Figure 36 Clothing Effects on the Feature Vector from the Bottom Half Mask (no DC component)

However, not every mask is suitable in these circumstances. Figure 37 illustrates the feature vector from the vertical line mask when the DC component is removed. The original sequence and the rain coat sequence have a similar shape, though it is doubtful they are similar enough for recognition to take place. The trench coat sequence has significantly less magnitude variation. This can be explained because the trench coat removes dynamical information from the sequence by obscuring the view of the legs.

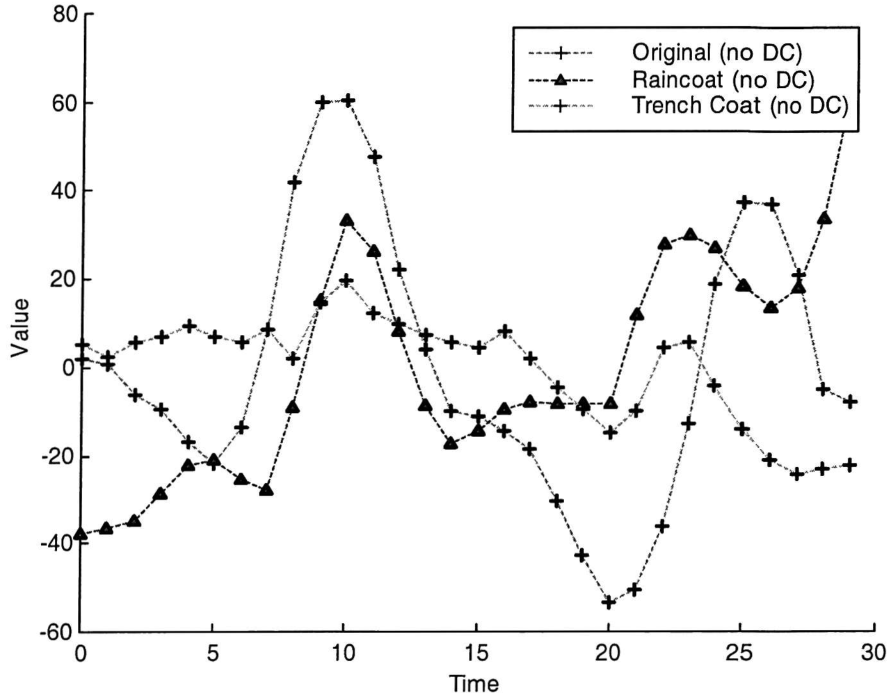


Figure 37 Clothing Effects on the Feature Vector from the Vertical Line Mask (no DC component)

Future work will concentrate on evaluating whether the recognition rates are affected when the subject is wearing different clothing than in the original database. We are unable to undertake these experiments at present as the database is still being processed.

9.2.2 Footwear

In the intra-subject variance database, subjects wore a variety of shoes including flip-flops, boots, trainers as well as their own shoes. How does changing the shoe of a subject affect their gait? Figure 38 illustrates the resulting feature vector using the vertical line mask when the subject is wearing a variety of footwear. With only a single subject it is difficult to discern any substantial differences due to footwear.

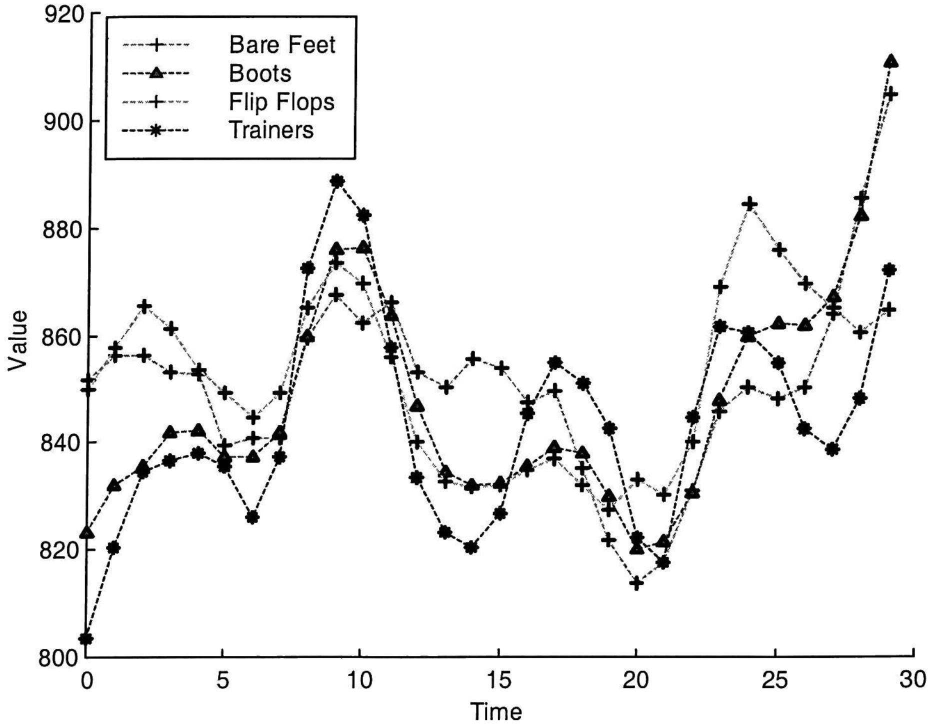


Figure 38 Does Footwear effect Gait? (Vertical Line Mask)

By focussing exclusively on the bottom half of the subject we gain further insight into how a change of footwear affects gait. Figure 39 illustrates the resulting plots from the bottom half mask for four different types of footwear. As can be seen there is minimal variation due to footwear. The slight changes between footwear require further investigation on the complete intra-subject variance database.

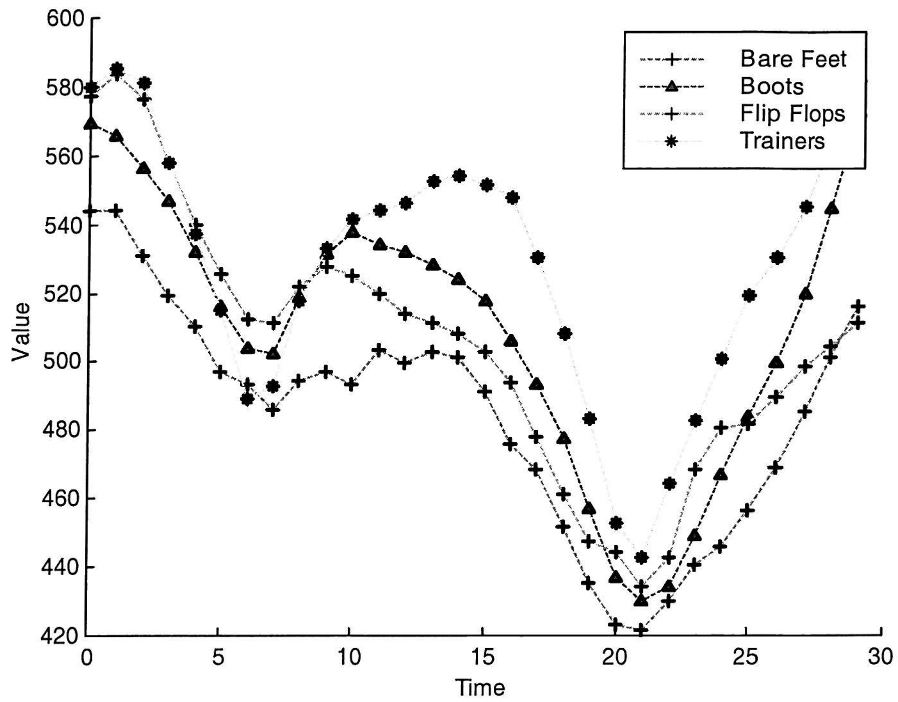


Figure 39 Does Footwear effect Gait? (Bottom Half Mask)

Further research is necessary to determine whether footwear does effect gait sufficiently for recognition to be a problem. When the intra-subject variance database is completed, experiments can be undertaken using normal shoes for each subject as the training database and using a variety of other shoes as the testing sequences.

9.2.3 Carrying Loads

In typical application, gait recognition algorithms need to be able to recognise a subject under a wide variety of conditions. The small intra-subject variance database films subjects carrying three different types of load: hand-bag, barrel bag and a rucksack. How does this affect gait and the resultant feature vectors? Clinical gait analysis and ergonomics communities have studied load carrying and concluded that when a subject is carrying an object they adjust the way they walk in order to minimise their energy expenditure [62, 63]. Previous work by Abdelkader and Davis [64] has focussed on detecting subjects carrying objects by observing distortions in the pendulum model of leg movement, but has not addressed the question of whether the change in gait pattern means that the subject can no longer be identified using gait recognition algorithms.

Carrying a load will affect the resultant feature vector in two different ways. Firstly, carrying a load is evidenced with changes in the silhouette, such as those illustrated in Figure 40. Secondly, the resultant feature vector will be changed since the subjects' gaits will adapt to load carrying.

(a) Hand bag (b) Shoulder Bag (c) Rucksack (d) Barrel Bag

Figure 40 Subjects carrying a variety of loads

If we considered both the spatial and temporal data there would be a clear difference between each subject, simply because a subject carrying a load consists of more pixels than a subject who does not. Figure 41 shows a plot using the vertical line mask (without the DC component) for the subject carrying no load and carrying a handbag, shoulder bag or rucksack.

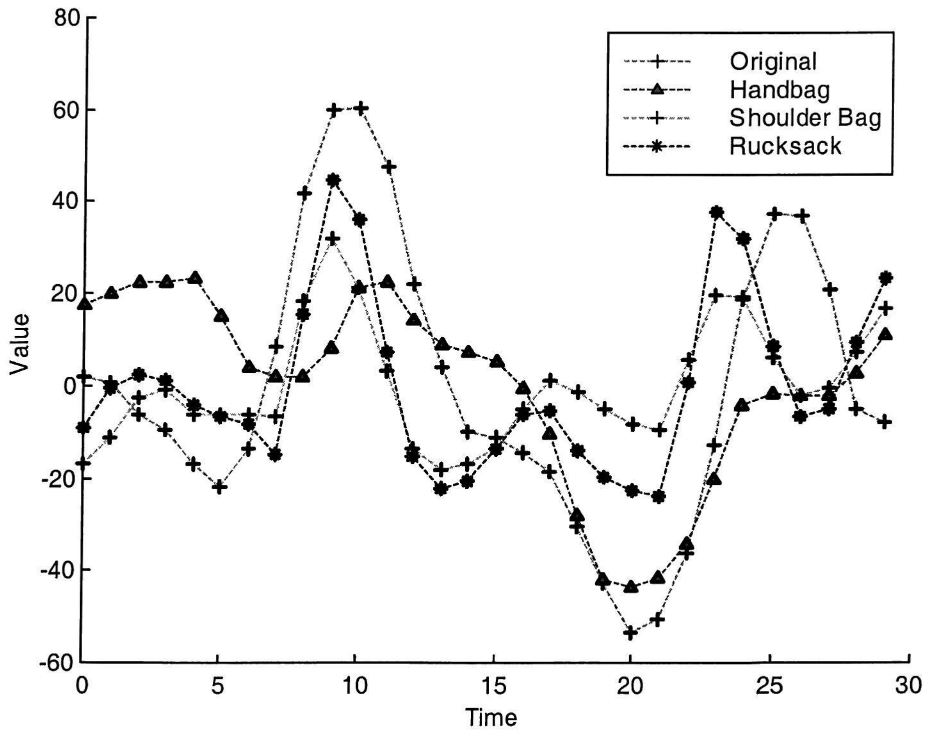


Figure 41 Area Mask data for Subject carrying various Loads

The graph once again illustrates the power of using only the temporal components of using gait. We repeated the experiment but this time using the bottom half mask (with no DC component). Figure 42 illustrates the resulting feature vectors.

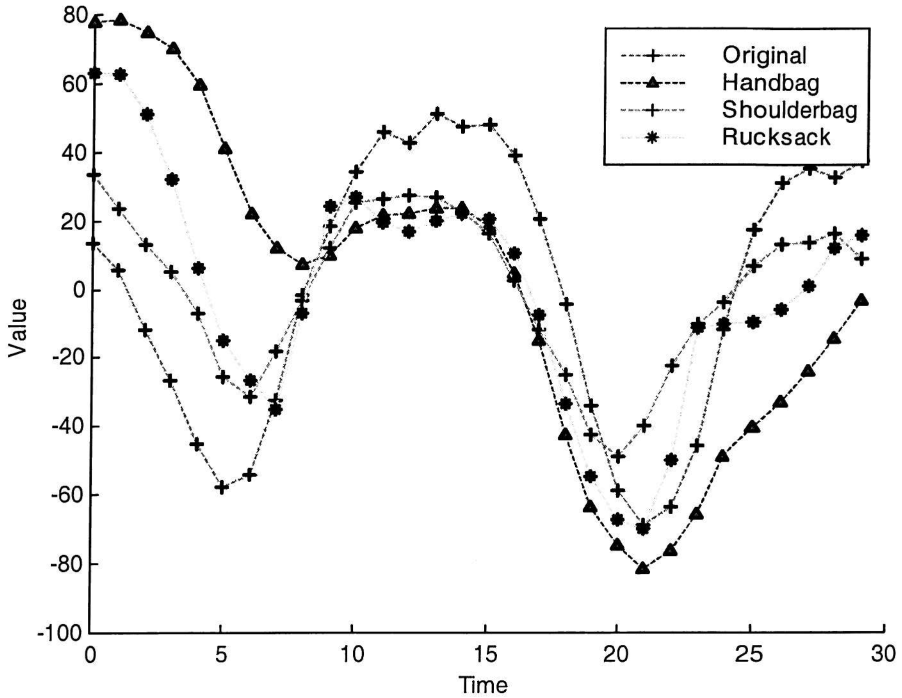


Figure 42 Bottom Half Mask (no DC) data for Subject carrying Various Loads

As can be seen the feature vectors are of similar shape, but differ in magnitude (due to the additional pixels in the silhouette because of load carrying). Further analysis of load carrying is necessary to determine whether the differences in magnitude would result in poor performance in typical application scenarios.

9.3 Discussion

This chapter has presented preliminary results from a new database which aims to address the area of inter-subject variance. Subjects are filmed under similar conditions to the SOTON database but are filmed with a variety of footwear, carrying loads and in different clothing.

Preliminary results have indicated that by using only the temporal components of gait (i.e. removing the DC components of the image sequences before applying area masks) the feature vectors remain similar even under adverse conditions, such as load carrying or wearing different clothes.

The results presented in this chapter come from a single sequence extracted from the intra-subject database. Further research will process the database in much greater detail and examine the effects of intra-subject variance on recognition performance.

Chapter 10 Conclusions and Future Work

10.1 Conclusions

In this thesis, we have presented two new contributions to automatic gait recognition based on recognising a subject by geometric properties of the silhouette's shape. The first approach, area masks, constructs a low dimensional signature which captures the dynamics of area change within specific portions of the silhouette sequence. The second approach, moment based descriptors, describes an image sequence in terms of geometric invariants. The SOTON database has been used to evaluate both techniques. The SOTON database is currently the largest of its kind, consisting of 114 subjects with at least eight samples of each subject.

Area masks have very low computational complexity and thus results can be calculated very quickly. Gait is different from other biometrics in that it is a function of both space and time and hence the dimensionality of the data is huge. By combining the results together from multiple area masks we can take advantage of this and capitalise on the tremendous amount of independent information available from each area mask, whilst still maintaining low computational complexity. Area masks achieve a recognition rate of between 30% and 43% when just a single mask is used. However, by combining seven area masks, a recognition rate of over 70% was achieved on the SOTON database. We have also shown that the majority of this recognition performance comes from the temporal components of gait, and have achieved a 50% recognition rate on the SOTON database by using the temporal components alone. This is an area which has not been addressed before and provides interesting results. In contrast, the recognition rate using the static components alone was 30%.

Previous research has indicated that the male and female walkers differ in terms of lateral body sway, with males tending to swing their shoulders from side to side more than their hips, and females tending to swing their hips more than their shoulders [10, 11]. This information is difficult to discern from the sagittal view used in the database. By using the mid-triangle mask (see Figure 5 (g)), area masks achieved a 77% correct discrimination rate between male and female subjects on the GEN-SOTON database. This is significantly better than chance (50%) and indicates that gender discrimination information is available in the dynamics of area change within that region of the silhouette.

Psychologists have typically assumed that gait is a symmetric pattern of movement [58, 59]. However recent research has suggested that limb dominance may mean that gait is not as symmetric as once thought [59]. Singh [60] and Du Chatinier et al [61] have also supported the notion of asymmetry in the normal population. To study the affect of symmetry on gait recognition we undertook experiments where the heel strikes and direction of walk was normalised. This is possible because the SOTON database consists of labelled subject data. We demonstrated that by using heel strike information and/or the direction of walk it was possible to increase the recognition rate of the system to over 80%. In addition, by considering the heel strike information, the performance of individual masks was also substantially improved. This research implies that the asymmetric part of gait is a good discriminator between subjects.

The part of the SOTON database used in this thesis has been designed to give almost perfect silhouettes to allow us to assess the basic properties of gait as a biometric. In real-world conditions silhouettes are likely to be far from perfect. To test the generalisation capabilities of area masks, we looked at four areas, noise analysis, mask centring, rotated silhouettes and rescaled images. To examine the affects of noise, we added salt and pepper noise in varying degrees to the original silhouettes. Performance, as expected, degraded as noise was added to the silhouettes. Although the noise simulated is not realistic, it does emphasise the problems that silhouette based techniques will face when presented with real-world data. Area masks rely on the subject being centred in a window. To simulate the affects of a misaligned window, we used a vertical line mask and shifted the position of the mask left or right by varying numbers of pixels. This showed, as expected, that the centring of the mask is crucial to the overall performance of the area based technique. Area masks are unable to deal with situations when the silhouette is rotated (as might occur due to poor camera calibration or due to real-world conditions).

One of the prime advantages of gait as a biometric is that it allows recognition from a distance. From a distance, the resolution of the human silhouette may be poor.

We rescaled the original 64 x 64 silhouettes to smaller sizes and tested the results. The results from this experiment were encouraging, as recognition rates did not drop significantly until the image size was less than a quarter of the original resolution. This is due to the large amount of dynamic information in a gait sequence that is visible, even at low resolutions.

We have reported on preliminary results from the intra-subject variance database which is presently under development at the University of Southampton. We have shown that by using the temporal components of gait we can compensate for the subject wearing different clothing or carrying loads. Further research needs to be undertaken to assess these properties empirically.

Moment based descriptors describe a gait sequence in terms of geometric invariants. We have shown that by using individual Hu invariant moments on the SOTON database recognition rates of between 22% and 44% could be achieved. By using a family of shape descriptors this can be greatly improved to almost 70%. The family of shape descriptors is formed by using a masking circle which determines portions of the silhouette to consider. This allows us to give a more complete description of the shape and results in better recognition rates. The performance level on the SOTON database by using a family of shape descriptors is very similar to that of area masks (approximately 70%).

Gender discrimination rates using moment based descriptors were similar to those using area masks. A correct gender discrimination rate of 75% was achieved on the GEN-SOTON database, which is almost ten percent less than that achieved using area masks. We showed that the Hu2 invariant discriminates better between genders than the Hu1 moment, although the Hu1 moment is a better person discriminator.

By considering the heel strike information present in the SOTON database, the recognition performance can be substantially increased, with recognition rates of over 80%. This once again demonstrates that the asymmetric component of gait provides much discrimination capability.

To simulate real world data we ran three sets of experiments for moment based descriptors. The first experiment analysed the effects of simulated noise (applied to the silhouettes) to recognition rates. The second experiment looked at low resolution silhouettes and examined at which point the silhouettes became too small to be able to discriminate between subjects, and the final experiment looked at the effect of rotation on the underlying feature vector. When noise was added to the silhouettes the recognition performance dropped accordingly. The results from scaling the silhouettes were more impressive with the performance of the system not degrading significantly until the images were of size 24 x 24 compared to the original 64 x 64 silhouettes. By using Hu moments the technique is approximately rotationally invariant and this was borne out in the experiments which showed that the feature vector does not change significantly when the silhouettes are rotated.

Our main contribution with this thesis has been to present two new techniques for automatic gait recognition. Both techniques achieve a recognition rate of approximately 70% on the SOTON database. We have shown that a recognition rate of over 50% is possible, even if no static cues are used. We have also demonstrated that the performance of both techniques can be significantly improved if heel strike information is taken into account. This demonstrates that gait, contrary to popular belief, does have an asymmetric component and that this component is a good discriminator between subjects. The remainder of this chapter describes future areas for research.

10.2 Future Work

Gait is a recent biometric, and we are only now beginning to see the construction of large databases with which to be able to evaluate the performance of techniques.

In evaluating the performance characteristics of area masks and moment based descriptors, it is clear that the silhouette extraction process plays a key part in the final recognition rates. As such, techniques for subject extraction need to be improved. The technique should be aimed at extracting human motion, rather than simply segmenting the background. A technique such as the velocity Hough transform [21] aims to extract a silhouette based on human movement patterns, rather than segmenting the background. Recent work by Lappas et al. [65, 66] has looked at improving this and developed a general purpose evidence based tracker. Following identification of a suitable technique to extract a subject from a background, a more realistic outdoor database can be constructed and used to evaluate both techniques in more detail.

Both techniques require that exactly one gait cycle of a subject is present. It is possible to extract exactly one cycle by identifying key frames in the image [20]. At present this is not implemented directly as part of the recognition procedure, as the SOTON database already consists of exactly one gait cycle of each subject.

We have also shown how it is possible to dramatically improve the recognition by considering the direction of travel and heel strike information. The recognition performance could be much improved if, as part of the silhouette extraction procedure, the subjects were labelled according to their heel strikes and direction of travel. We have shown that even by simply considering the direction of travel, which is easily extracted from a gait sequence, performance can be increased substantially. This does, however, rely on large amounts of training data being present with the subject walking in both directions.

The classifier chosen for this system was a simple nearest neighbour rule in the feature space. By using a different classifier it may be possible to improve the recognition rates. We have examined using canonical analysis as a classifier, and rejected this because of poor performance and computational complexity. However, there are a wide variety of classifiers available such as support vector machines, kernel discriminant analysis and neural networks all of which should be evaluated to see if they can provide better performance than using a simple nearest neighbour classifier.

To address the subject of how much an individual subject's gait varies, further analysis of the intra-subject variance database needs to be undertaken. Using this database, we can examine the effects of footwear on gait, together with issues such as clothing and load carrying. Preliminary studies (Chapter 9) have shown that by utilising the temporal component of gait we can gain tolerance to load carrying and clothing conditions, but as we have shown in Chapter 5 this comes at the expense of recognition performance.

In this thesis we have only considered a subject walking normal to the camera's plane of view. For gait recognition to be achievable this condition needs to be relaxed and we need to consider subjects walking at an angle to the camera. Analysis of the simple pendulum shown in section 3.5 suggests that we will be able to adapt the size of the area mask to compensate for small angles of rotation.

Further research is needed to determine whether the moment based descriptor technique can be adapted to cope with the subject walking at an angle to the camera. At high angles (when the subject is walking almost towards the camera) it is doubtful that there exists enough dynamic cues to be able to discriminate between subjects, though further research would be necessary to prove this. In addition to the subject walking at an angle to the camera we should evaluate what happens to gait when the subject is walking on an incline.

The moment based descriptors technique still has plenty of scope for improvement. An investigation needs to be undertaken to see what happens when different masking shapes other than circles are used. In this respect, masks used in the area mask technique could be used and the affects on database performance analysed. This would give moment based descriptors the specificity of area mask descriptors and hopefully improve the performance of the algorithm on large databases.

Clearly one of the most pressing concerns with gait as a biometric is the large variety of circumstances it can be found. To fully evaluate gait as a biometric requires the construction of several large databases filmed under a variety of conditions. There are many variables to consider such as load carrying, angle to camera, clothing etc. This thesis has presented the first large scale evaluation of gait as a biometric under controlled conditions. Future work will concentrate on seeing whether these techniques can be extended to cope with a greater variety of circumstances.

Appendix A – Database Overview

The table below shows an overview of the SOTON database. The number below each silhouette refers to the label used to uniquely identify each subject. Numbers 82 and 93 are missing from the database because insufficient numbers of sequences were captured for both subjects.

The SOTON database contains all 114 subjects and eight sequences of each subject were used (912 sequences in total). The gender discrimination database, GEN-SOTON consists of only 48 sequences with 24 of each gender. The female subjects used in the GEN-SOTON database were 2, 31, 33, 35, 36, 37, 39, 41, 49, 57, 67, 73, 81, 82, 90, 96, 97, 98, 99, 101, 110, 114, 115, and 116. The male subjects were chosen at random and were 1, 15, 16, 18, 19, 20, 23, 27, 32, 40, 45, 48, 55, 64, 78, 85, 86, 91, 95, 102, 103, 105, 107, and 109.

					
1	2	3	4	5	6
					
7	8	9	10	11	12
					
13	14	15	16	17	18
					
19	20	21	22	23	24

25 26 27 28 29 30

31 32 33 34 35 36

37 38 39 40 41 42

43 44 45 46 47 48

49 50 51 52 53 54

55 56 57 58 59 60

61 62 63 64 65 66

67 68 69 70 71 72

73 74 75 76 77 78

79 80 81 83 84 85

86 87 88 89 90 91

92 94 95 96 97 98

99 100 101 102 103 104

105 106 107 108 109 110

111 112 113 114 115 116

Figure 43 Database Overview

The following sequence of images shows a sample gait sequence from subject 46. Each gait sequence consists of a complete gait cycle which is equivalent to two steps.

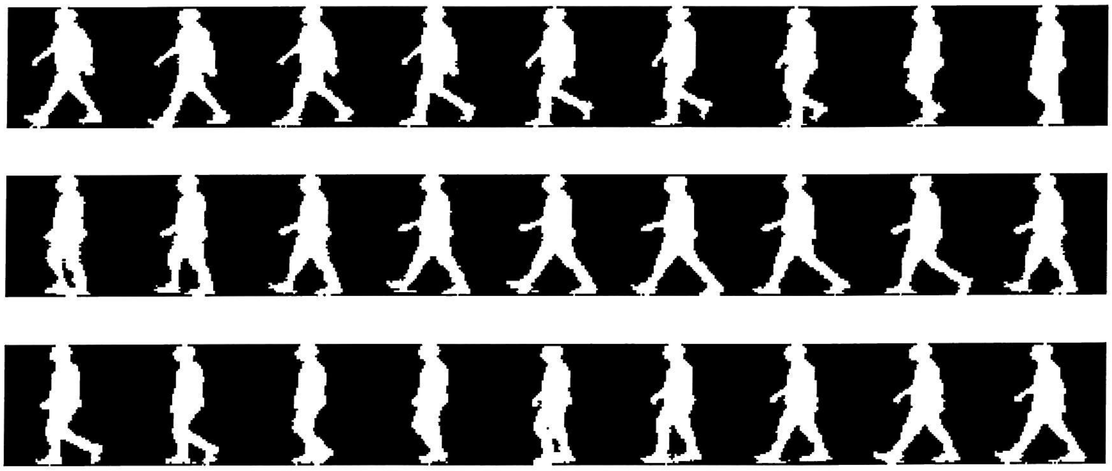


Figure 44 Sample Gait Sequence

List of publications

- J. Foster, M. Nixon, and A. Prügel-Bennett, "New Area Based Metrics for Automatic Gait Recognition," *Proc. British Machine Vision Conference 2001*, pp. 233-242, 2001.
- J. Foster, M. Nixon, and A. Prügel-Bennett, "New Area Measures for Automatic Gait Recognition," *BMVA Workshop Understanding Visual Behaviour*, 2001.
- J. Foster, M. S. Nixon, and A. Prügel-Bennett, "New Area Based Measures for Gait Recognition," *Proc. Audio- and Video-Based Biometric Person Authentication*, pp. 312-7, 2001.
- J. Foster, M. S. Nixon and A. Prügel-Bennett, "Automatic Gait Recognition using Area Based Metrics", submitted to Pattern Recognition Letters.
- J. Foster, M. S. Nixon and A. Prügel-Bennett, "Automatic Gait Recognition via Moment Based Descriptors", submitted to Soft Computing 2002.

References

[1] J. Shutler and M. Grant, "The Large Southampton Human ID Gait Database," *Technical Report, ISIS Research Group, ECS, University of Southampton*, 2002.

[2] G. Harris and P. Smith, *Human Motion Analysis*: IEEE Press, 1997.

[3] A. Kairento and G. Hellen, "Biomechanical Analysis of Walking," *Biomechanics*, vol. 14, pp. 671-678, 1981.

[4] J. Perry, *Gait Analysis - Normal and Pathological Function*: Slack, 1992.

[5] S. V. Stevenage, M. S. Nixon, and K. Vince, "Visual Analysis of Gait as a Cue to Identity," *Applied Cognitive Psychology*, pp. 513-526, 1999.

[6] M. P. Murray, "Gait as a Total Pattern of Movement," *American Journal of Physical Medicine*, vol. 46, pp. 290-332, 1967.

[7] M. P. Murray, A. B. Drought, and R. C. Kory, "Walking Patterns of Normal Men," *Journal of Bone Joint Surgery*, vol. 46-A, pp. 335-360, 1964.

[8] G. Johansson, "Visual Perception of Biological Motion and a Model for its Analysis," *Perception Psychophysics*, vol. 14, pp. 201-211, 1973.

[9] J. E. Cutting and L. T. Kozlowski, "Recognizing Friends by their Walk: Gait Perception Without Familiarity Cues," *Bulletin of the Psychonomic Society*, vol. 9, pp. 353-356, 1977.

[10] L. T. Kozlowski and J. E. Cutting, "Recognizing the Sex of a Walker from a Dynamic Point-Light Display," *Perception and Psychophysics*, vol. 21, pp. 575-580, 1977.

[11] G. Mather and L. Murdoch, "Gender Discrimination in Biological Motion Displays based on Dynamic Cues," *Proc. Royal Society*, pp. 273-279, 1994.

[12] A. Bobick and A. Johnson, "Gait Recognition using Static, Activity-specific Parameters," *IEEE Computer Vision and Pattern Recognition*, vol. 1, pp. 423-430, 2001.

[13] H. Murase and R. Sakai, "Moving Object Recognition in Eigenspace Representation: Gait Analysis and Lip Reading," *Pattern Recognition Letters*, vol. 17, pp. 155-162, 1996.

[14] J. J. Little and J. E. Boyd, "Recognising People by Their Gait: the Shape of Motion," *Videre, International Journal of Computer Vision*, vol. 14, pp. 83-105, 1998.

[15] M. S. Nixon, J. N. Carter, D. Cunado, P. S. Huang, and S. V. Stevenage, "Automatic Gait Recognition," in *BIOMETRICS - Personal Identification in Networked Society*, A. Jain, R. Bolle, and S. Pankanti, Eds., 1999, pp. 231-249.

[16] S. A. Niyogi and E. H. Adelson, "Analyzing and Recognizing Walking Figures in XYT," *Proc. Computer Vision and Pattern Recognition*, pp. 469-474, 1994.

[17] C. BenAbdelkader, R. Cutler, and L. Davis, "Stride and Cadence as a Biometric in Automatic Person Identification," *Proc. 5th IEEE Face and Gesture Recognition*, pp. 372-378, 2002.

[18] C. BenAbdelkader, R. Cutler, H. Nanda, and L. Davis, "EigenGait: Motion-Based Recognition Using Image Self-Similarity," *Proc. 3rd International Conference AVBPA*, pp. 289-294, 2001.

[19] C. BenAbdelkader, J. Cutting, and L. Davis, "Motion-based Recognition of People in EigenGait Space," *Proc. 5th IEEE Face and Gesture Recognition*, pp. 378-384, 2002.

[20] R. Collins, R. Gross, and J. Shi, "Silhouette-based Human Identification from Body Shape and Gait," *Proc. 5th IEEE Face and Gesture Recognition*, pp. 366-371, 2002.

[21] D. Cunado, J. M. Nash, M. S. Nixon, and J. N. Carter, "Gait Extraction and Description by Evidence-Gathering," *Proc. Audio and Video based Biometric Person Authentication*, pp. 43-48, 1999.

[22] D. Cunado, M. S. Nixon, and J. N. Carter, "Automatic Gait Recognition via Model-Based Evidence Gathering," *Proc. AutoID99: IEEE Workshop on Automated ID Technologies*, pp. 27-30, 1999.

[23] J. Foster, M. Nixon, and A. Prugel-Bennett, "New Area Based Metrics for Automatic Gait Recognition," *Proc. British Machine Vision Conference 2001*, pp. 233-242, 2001.

[24] J. P. Foster, M. S. Nixon, and A. Prugel-Bennett, "New Area Based Measures for Gait Recognition," *Proc. Audio- and Video-Based Biometric Person Authentication*, pp. 312-7, 2001.

[25] J. B. Hayfron-Acquah, M. S. Nixon, and J. N. Carter, "Recognising Human and Animal Movement by Symmetry," *Proc. IEEE International Conference on Image Processing*, pp. 290-293, 2001.

[26] J. B. Hayfron-Acquah, M. S. Nixon, and J. N. Carter, "Automatic Gait Recognition by Symmetry Analysis," *Proc. Audio-and-Video-Based Biometric Person Authentication*, pp. 272-277, 2001.

[27] P. S. Huang, C. J. Harris, and M. S. Nixon, "Human Gait Recognition in Canonical Space using Temporal Template," *IEE Proc. Vision Image and Signal Processing*, vol. 146, pp. 93-100, 1999.

[28] P. S. Huang, C. J. Harris, and M. S. Nixon, "Recognising Humans by Gait via Parametric Canonical Space," *Artificial Intelligence in Engineering*, vol. 13, pp. 359-366, 1999.

[29] A. Y. Johnson and A. F. Bobick, "A Multi-View Method for Gait Recognition Using Static Body Parameters," *3rd International Conference AVBPA*, pp. 301-311, 2001.

[30] L. Lee and W. Grimson, "Gait Analysis for Recognition and Classification," *Proc. 5th IEEE Face and Gesture Recognition*, pp. 155-162, 2002.

[31] P. J. Phillips, S. Sarkar, I. Robledo, P. Grother, and K. Bowyer, "Baseline Results for the Challenge Problem of Human ID Using Gait Analysis," *Proc 5th Automatic Face and Gesture Recognition*, pp. 137-143, 2002.

[32] J. Shutler, M. Nixon, and C. Harris, "Zernike Velocity Moments for Description and Recognition of Moving Shapes," *British Machine Vision Conference 2001*, pp. 705-714, 2001.

[33] J. D. Shutler, M. S. Nixon, and C. J. Harris, "Statistical Gait Recognition via Temporal Moments," *4th IEEE Southwest Symposium on Image Analysis and Interpretation, Texas*, pp. 291-295, 2000.

[34] J. D. Shutler, M. S. Nixon, and C. J. Harris, "Statistical Gait Recognition via Velocity Moments," *IEE Electronics and Communications, Colloquium, Visual Biometrics, London*, pp. 1-5, 2000.

[35] C.-Y. Yam, M. S. Nixon, and J. N. Carter, "Extended Model-Based Automatic Gait Recognition of Walking and Running," *Proc. 3rd Audio- and Video-Based Biometric Person Authentication*, pp. 284-294, 2001.

[36] C.-Y. Yam, M. S. Nixon, and J. N. Carter, "Gait Recognition by Walking and Running: A Model-based Approach," *Proc. 5th Asian Conference on Computer Vision*, pp. 1-6, 2002.

[37] B. Bhanu and J. Han, "Individual Recognition by Kinematic-Based Gait Analysis," *Proc. 16th International Conference on Pattern Recognition*, vol. 3, pp. 343-347, 2002.

[38] I. Vega and S. Sarkar, "Experiments on Gait Analysis by Exploiting the Nonstationarity in the Distribution of Feature Relationships," *Proc. 16th International Conference on Pattern Recognition*, vol. 1, pp. 1-5, 2002.

[39] E. Tassone, G. West, and S. Venkatesh, "Temporal PDMs for Gait Classification," *Proc. 16th International Conference on Pattern Recognition*, vol. 2, pp. 1065-1069, 2002.

[40] L. Wang, W. Hu, and T. Tan, "A New Attempt to Gait-Based Human Identification," *Proc. 16th International Conference on Pattern Recognition*, vol. 1, pp. 115-119, 2002.

[41] J. Nash, J. N. Carter, and M. Nixon, "Dynamic Feature Extraction via the Velocity Hough Transform," *Pattern Recognition Letters*, vol. 18, pp. 1035-1047, 1997.

[42] J. Hayfron-Acquah, M. Nixon, and J. Carter, "Human Identification by Spatio-Temporal Symmetry," *Proc. 16th International Conference on Pattern Recognition*, vol. 1, pp. 632-636, 2002.

[43] H. Murakami and V. Kumar, "Efficient Calculation of Primary Images from a set of Images," *IEEE Transactions on Pattern analysis and Machine Intelligence*, vol. 4, pp. 511-515, 1982.

[44] R. A. Fisher, "The Statistical Utilization of Multiple Measurements," *Annals of Eugenics*, vol. 8, pp. 376-386, 1938.

[45] D. L. Swets and J. J. Weng, "Using Discriminant Eigenfeatures for Image Retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, pp. 831-836, 1996.

[46] P. Hall, D. Marshall, and R. Martin, "Incremental Eigenanalysis for Classification," *Proc. British Machine Vision Conference*, pp. 286-295, 1998.

[47] I. J. Schoenberg, "Contribution to the Problem of Approximation of Equidistant Data by Analytic Functions," *Quart. Appl. Math*, vol. 4, pp. 45-99,112-141, 1946.

[48] J. N. Carter and M. S. Nixon, "On measuring Gait Signatures which are Invariant to their Trajectory," *Measurement and Control*, vol. 32, pp. 265-269, 1999.

[49] A. Sluzek, "Identification and Inspection of 2-D objects using New Moment-Based Shape Descriptors," *Pattern Recognition Letters*, vol. 16, pp. 687-697, 1995.

[50] M. Hu, "Visual Pattern Recognition by Moment Invariants," *IRE Trans. Inform. Theory*, vol. 8, pp. 179-187, 1962.

[51] D. Cyganski, J. Orr, and Z. Pinjo, "A Tensor Operator Method for Identifying the Affine Transformation Relating Image Pairs," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 351-363, 1983.

[52] J. Sun, M. Walters, N. Svensson, and D. Lloyd, "The Influence of Surface Slope on Human Gait Characteristics: A study of Urban Pedestrians Walking on an Inclined Surface," *Ergonomics*, vol. 39 (4), pp. 677-692, 1996.

[53] A. Webb, *Statistical Pattern Recognition*: Arnold, 1999.

[54] P. Phillips, S. Sarkar, I. Robledo, P. Grother, and K. Bowyer, "The Gait Identification Challenge Problem: Data Sets and Baseline Algorithm," *Proc. 16th International Conference on Pattern Recognition*, vol. 1, pp. 385-389, 2002.

[55] K. Soudan, "Standardization of Gait Kinematic data using a Gait Symmetry Index and Fourier analysis," *In: Biomechanics: Principles and Applications, Development in Biomechanics*, vol. 1, pp. 135-140, 1982.

[56] W. Herzog, B. Nigg, L. Read, and E. Olsson, "Asymmetries in Ground Reaction Force Patterns in Normal Human Gait," *Medicine in Science and Sports and Exercise*, vol. 21, pp. 110-114, 1982.

[57] C. Gabbard, "Coming to Terms with Laterality," *Journal of Psychology*, vol. 131, pp. 561-564, 1997.

[58] J. T. Cutting, D. R. Proffitt, and L. T. Kozlowski, "A Biomechanical Invariant for Gait Perception," *Journal of Experimental Psychology: Human Perception and Performance*, pp. 357-372, 1978.

[59] H. Sadeghi, P. Allard, F. Prince, and H. Labelle, "Symmetry and Limb Dominance in Able Bodied Gait: A Review," *Gait and Posture*, vol. 12, pp. 34-45, 2000.

[60] I. Singh, "Function Asymmetry in the Lower Limbs," *Acta Anatomica*, vol. 77, pp. 131-138, 1970.

[61] K. DuChatinier and R. Rozendal, "Temporal Symmetry Gait of Selected Normal Subjects," *Anatomy*, vol. 73, 1970.

[62] P. Martin and R. Nelson, "The Effect of Carried Loads on the Walking Patterns of Men and Women," *Ergonomics*, vol. 29(10), 1986.

[63] H. Kinoshita, "Effects of Different Loads and Carrying Systems on Selected Biomechanical Parameters describing Walking Gait," *Ergonomics*, vol. 28(9), 1985.

[64] C. BenAbdelkader and L. Davis, "Detection of People Carrying Objects: a Motion-based Recognition Approach," *Proc. 5th IEEE Face and Gesture Recognition*, vol. 1, pp. 378-383, 2002.

[65] P. Lappas, J. N. Carter, and R. Damper, "Object Tracking via the Dynamic Velocity Hough Transform," *Proc IEEE International Conference on Image Processing*, pp. 3710374, 2001.

[66] P. Lappas, J. N. Carter, and R. Damper, "Robust Evidence-based Object Tracking," *Pattern Recognition Letters*, vol. 23, pp. 253-260, 2002.