VELOCITY MOMENTS FOR HOLISTIC SHAPE DESCRIPTION OF TEMPORAL FEATURES

By
Jamie D. Shutler
B.Eng.(Hons)

A thesis submitted for the degree of
Doctor of Philosophy

Department of Electronics and Computer Science,
University of Southampton,
United Kingdom.

July 2002

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING

ELECTRONICS AND COMPUTER SCIENCE DEPARTMENT

Doctor of Philosophy

Velocity Moments for Holistic Shape Description of Temporal Features

by Jamie D. Shutler

The increasing interest in processing sequences of images (rather than single ones) motivates development of techniques for sequence-based object analysis and description. Accordingly, new velocity moments have been developed to describe an object, not only by its shape but also by its motion through an image sequence. These moments are an extended form of centralised moments and compute statistical descriptions of the object and its behaviour. Two variations of this new technique are presented. The first uses the non-orthogonal Cartesian basis, while the second utilises the orthogonal Zernike one. Despite their difference in basis, both techniques exhibit favourable characteristics. Evaluation illustrates the advantages of using a complete image sequence (over single images), exploiting temporal correlation to improve a shape's statistical description, while also improving the performance of these statistical features under less favourable application scenarios, including occlusion and noise. To further characterise the velocity moments, they have been applied to gait recognition - a potential new biometric. Good recognition results have been achieved using relatively few features and basic feature selection and classification techniques. However, the prime aim of this new technique is to allow the generation of statistical features which encode shape and motion information, with generic application capability. Theoretical and applied analyses show the potential of this new sequence-based statistical technique and highlight the consistency of its performance attributes with those of conventional moments.

Contents

List of Figures

Nomenclature

η_{pq}	Two dimensional scale-normalised centralised moment
μ_{pq}	Two dimensional centralised moment
$\overline{x}$	x axis centre of mass
$\overline{y}$	y axis centre of mass
θ	Polar co-ordinate angle
$A_{pq\mu\gamma}$	Zernike velocity moment
A_{pq}	Two dimensional Zernike moment
F	Fisher statistic
$f(x)$	One dimensional continuous function
$f(x,y)$	Two dimensional continuous function
F_s	Scheffe post-hoc test value
F_{crit}	Fisher statistic critical value
I	Number of images
I_n	nth Hu invariant moment
k-nn	k-nearest neighbour classifier
m_{pq}	Two dimensional Cartesian moment
$P_{i_{xy}}$	ith image discrete pixel P_{xy}
P_{xy}	Discrete pixel
r	Polar co-ordinate radius
$R_{mn}(r)$	Orthogonal radial polynomial
$S(i,p,q)$	ith velocity moment spatial expression
$U(i,\mu,\gamma)$	ith velocity moment motion expression
$V_{mn}(x,y)$	Zernike polynomial
$vm_{pq\mu\gamma}$	Cartesian velocity moment
ANOVA	ANalysis Of VAriance
COM	Centre of mass
ST	Spatial template (binary silhouette)
TT	Temporal template (optical flow)

Acknowledgements

I would like to thank my supervisor, Mark Nixon for his continuous help and guidance throughout the duration of this research, and for his endless chats and stories about beer and hazardous activities. I'd like to thank all my friends at work (Mike, Karl, Richard, Jun, Chew-yean and Jaz), who provided endless humorous distractions! - especially Mike who even managed to read this wedge and apologies to Jaz, as I think Mike has won! My thanks must also go to my eldest sister, Karen, without whom I probably would have never even gone to university. Nice one! A special thank-you to all my great friends, both present and lost, whom helped me realise what life was really all about, especially Dan 'the kite surfing man', Kev, Aidan, Dave, my brother Paul for all his great parties and my 'pressure-point-prodding' :o) sister Martine. Gjobmmz, b tqfdjbm uibol-zpv up Lbsfo, gps bmm ifs mpwf boe ibqqjoftt jo uif zfbst mfbejoh vq up nf xsjujoh vq. And if anyone hears me talking about 'furthering my education', then bop me over the head with a saucepan, quickly!

I would like to dedicate this work to my Mum and Dad, who have always been there for me.

Have fun!

Majic!

THE END!

☺

Chapter 1

Context and contributions

1.1 Motivation

Traditionally, image processing and computer vision has largely focussed on analysing single images for their content. With the introduction of cheaper camera systems, along with digital video, the need and interest in processing streams of video footage continues to increase. One way in which this can be achieved is simply to separately analyse each consecutive image in the sequence. This produces frame orientated results that hold no information about how the images within the sequence relate to each other. More recent approaches use temporal information from within the image sequence to their advantage, exploiting the correlation between images within the sequence. For example, the popular MPEG [52] compression technique for video sequences updates the motion information in consecutive images, reducing the bandwidth by not updating static objects and therefore exploiting the high correlation between images.

One large area of image processing and computer vision is pattern classification. In general this area is divided into model-based and statistical methods. A few recent model-based methods have utilised the inclusion of temporal information into the descriptor, including the velocity Hough transform (VHT) [56], XYT slices to recognise human articulated motion [59] and velocity snakes [63, 67]. These techniques locate and describe a shape using both its motion and boundary description. Alternatively shapes can be described by their structure or distribution, using statistical techniques. However, the inclusion of temporal information and correlation into these statistical techniques is relatively unexplored. For example, Little [46] linked pairs of images using statistical moments, while Rosales [68] encoded an image sequence into a single image, and then described the single image by moments. However, these techniques do not consider the image sequence as a single entity (within the descriptor), in contrast with the previously mentioned model based methods.

Here we are interested in statistically describing a time varying, or temporal sequence of images. The image sequences analysed consist of either a static or a deforming shape. Unlike the statistical methods mentioned above, we are primarily interested in describing the sequence as a whole, rather than describing each separate image within the sequence. This enables us to exploit temporal correlation to provide an improved shape descriptor from a sequence of images. It is important to note that achieving high classification rates requires some application-specific selection of features and, as such, our primary focus (from a classification point of view) is in enabling generic selection of appropriate features. The generated features are descriptors encoding both shape and motion, thus enabling description, analysis and classification.

1.2 Temporal correlation

The analysis of an image sequence, in contrast to analysing a single image allows the exploitation of the temporal statistics of the sequence. The technique of using temporal correlation within an image sequence has previously been used to enhance feature extraction and description methods, examples include the previously mentioned VHT [56] and velocity snakes [67]. The use of an image sequence can be used to enrich a shape description, providing further information about the shape's structure, while also allowing its motion to be studied. For example, multiple images of a simple object, under noise conditions can be used to produce a single reduced-noise (averaged) image of the object. Figures 1.1, 1.2 and 1.3 demonstrate this. Additive zero mean ($\mu = 0$), unit variance ($\sigma^2 = 1$) random Gaussian noise has been added to a simple 100×100 binary image of the shape shown in Figure 1.1(a). Figure 1.1(b) shows the result of adding the noise to this single image. The shape is barely visible, while the contour plot of the grey-level values in Figure 1.1(c) shows considerable confusion between the foreground shape and the noisy background. Accumulating and averaging the pixel values over ten images (of the same original shape but with different random noise) produces the results in Figure 1.2. Here the shape is becoming clearer, separating from the background noise and is visible in the contour plot of Figure 1.2(b). The perimeter of the shape is beginning to appear on the projection displayed at the base of the plot. Continuing the process and accumulating over forty images produces the results in Figure 1.3. Here the noise floor level has dropped (in comparison to Figure 1.2(b)), while the foreground shape has become more prominent, producing an increased signal to noise ratio. Accumulating information using a sequence of images, instead of analysing a single image has improved the shape's contrast in the presence of Gaussian noise, as shown by comparing Figure 1.1(b) and Figure 1.3(a). More formally, the error

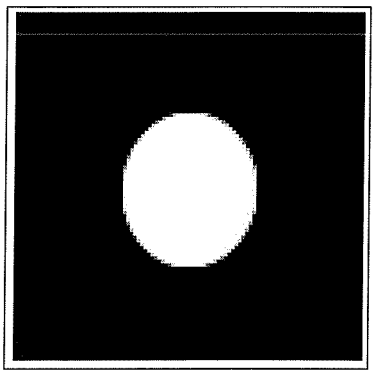
in the estimate of the average (image) is $O(\frac{1}{N})$, where N is the number of samples (or images). Therefore as $N \rightarrow \infty$ the error approaches zero.

1.3 Shape description for classification

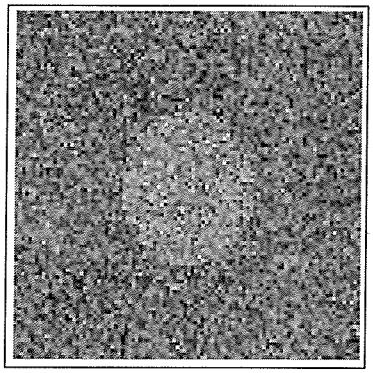
Once a shape has been isolated in a scene (by any number of feature extraction methods), it may be desirable to describe it, for classification. Shape description can use either *region* or *boundary* information. The former describes the shape with respect to its structure, while the latter is a representation of the shape's perimeter. It is often desirable for these shape descriptors to possess selected invariance properties. These include scale, rotation, position and even perspective invariance, allowing the shape description to be more versatile. First we discuss the boundary case, and then move on to regions.

Boundary descriptors can be efficient in terms of computing resources, as only a small proportion of the overall image is processed. Simple chain codes [21] describe a boundary by defining each pixel's location in terms of its previous neighbour, while traversing the contour. Fourier descriptors [86] describe the contour by a series of closed curves, a variation of the standard Fourier series time domain analysis. This enables contours to be described in a continuous manner (allowing reduced error when scaling). There exist many variations of Fourier descriptors eg. [23]. Alternatively, the curvature scale space [51] of a shape can be used as a boundary descriptor. Here the points of inflection, found as a shape's contour is traversed, are used as features to describe the contour. One problem with these boundary descriptions is that any important descriptive information held within the boundary is lost.

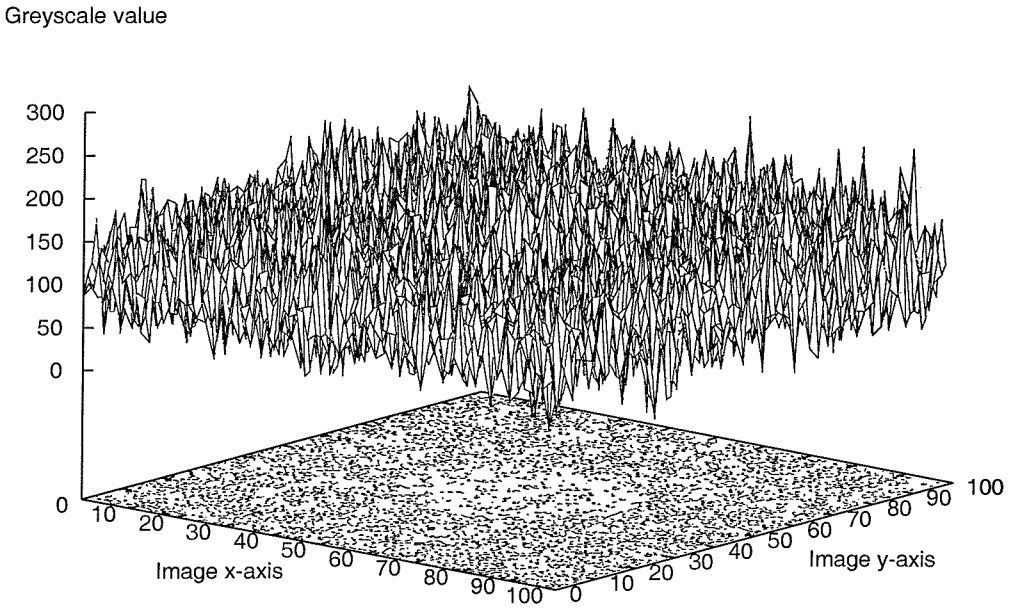
Shape descriptions via skeletons, first proposed by Blum [7] use the medial axis function (however, further variations exist). In this paper, Blum describes a shape's boundary in terms of its interaction with a series of energy functions which propagate through the interior of a binary shape. This is a compacted form of region description. Various simple *region* descriptors exist [58] including measures of area, compactness (or roundness - the ratio of perimeter to area) and elongation. A shape's topology can be described, where the interest is in properties of the region that do not change, even if the shape does - although this excludes tearing of the shape, or joining it with another. Convex hulls can be used to measure the number or size of any concavities within a shape, perhaps determining the differences between the letters 'O' and 'C'. Fourier descriptors can also be used for region description [39]. Extremal points in their simplest form can be used to determine the bounding area of a shape. By connecting them together the major and minor axis of a shape can be determined. These axes in turn can be used to define characteristics like the aspect ratio of the object.



(a) Original image.

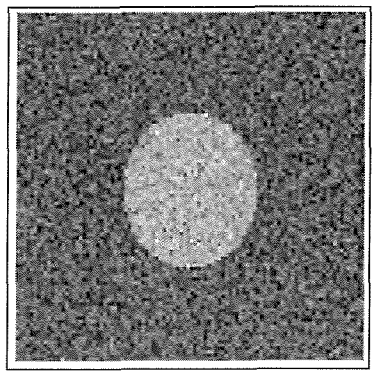


(b) Additive Gaussian noise.

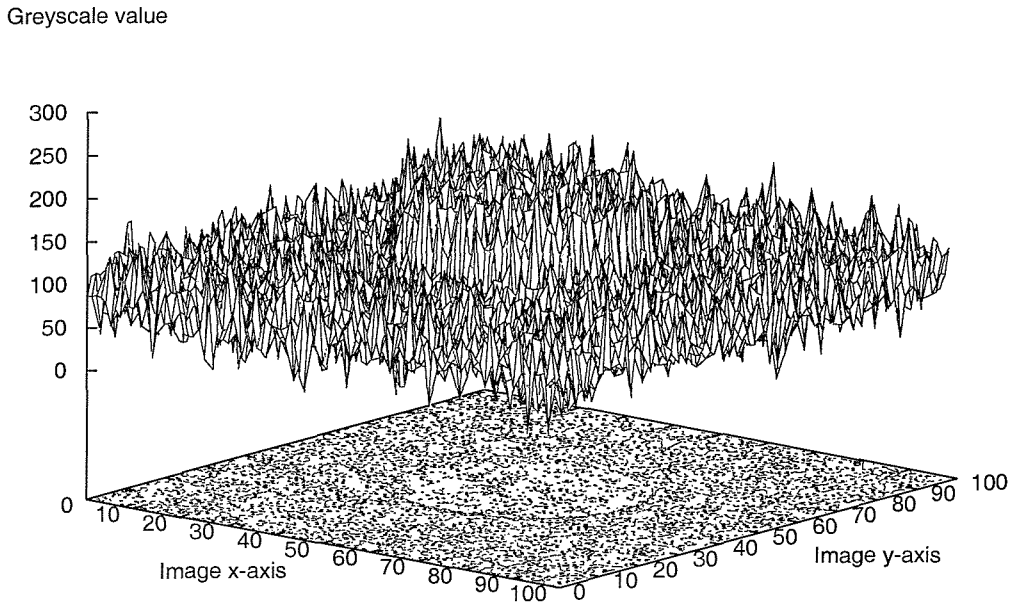


(c) Surface contour plot

Figure 1.1: The original image, the image with additive Gaussian noise and the corresponding surface plot of the image.

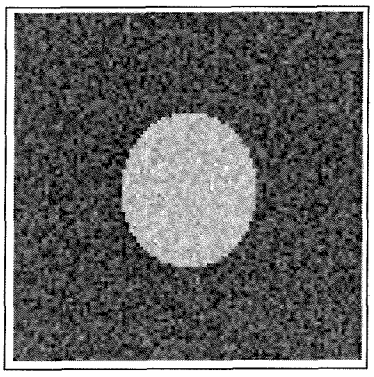


(a) Additive Gaussian noise.

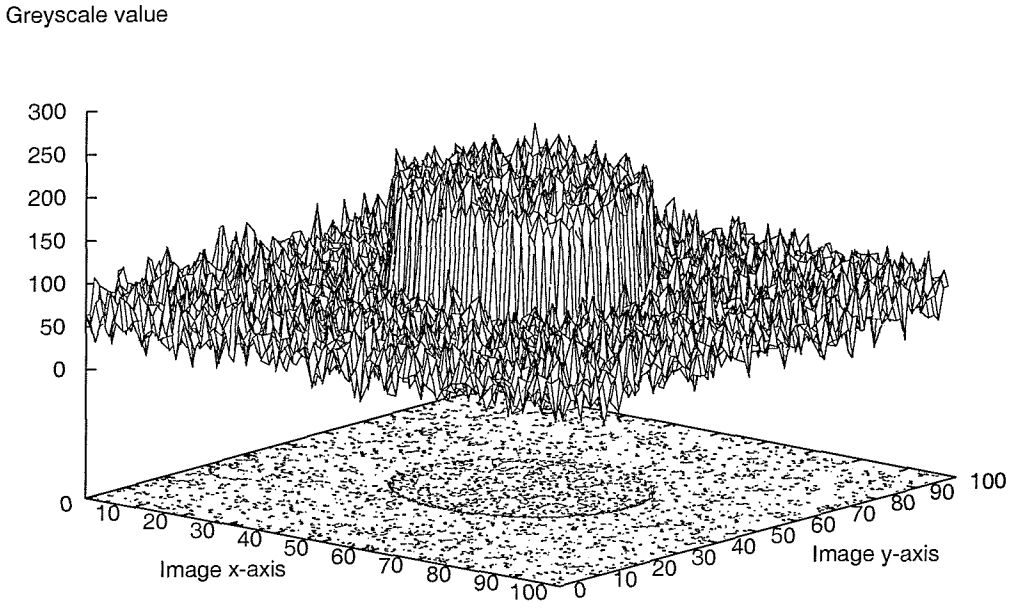


(b) Surface contour plot

Figure 1.2: The accumulated result of 10 images with additive Gaussian noise and the corresponding surface plot of the image.



(a) Additive Gaussian noise.



(b) Surface contour plot

Figure 1.3: The accumulated result of 40 images with additive Gaussian noise and the corresponding surface plot of the image.

Many of the global region descriptions (including area, spread, kurtosis and axes' orientation) can also be determined through a general framework called statistical moments. These global descriptions collate low frequency information about the image. In addition, the more complicated high frequency image content is available through the same framework. In this thesis we are primarily interested in statistical moments, as applied to images [27]. They describe a region, or shape in terms of its distribution. Depending on the particular type of moments used, various invariance properties are available (including rotation, scale and position).

1.4 Statistical moments

Statistical moments extract information from a signal, describing its distribution and make-up. Simple properties can be extracted including area (mass), mean and variance, while more specific information is available, i.e. projection information including kurtosis, skewness, spread and radii of gyration. The application of classical moments to two-dimensional images was first considered in the early sixties by Hu [27, 28]. Hu tested their validity using a simple experiment to recognise written characters. Hu was only concerned with images without noise, but further work by Teh [80] showed that traditional moment performance degrades when the view is occluded or noisy.

A survey of moment based techniques with respect to computer vision, Prokop [65], details many of the current techniques regarding representation and recognition. Belkasim [3] presents a comparative study of moment invariants, while Mukundan [53] provides descriptions of most of the current moment techniques, along with background information and applications. In general the different types of moments fall into two categories, orthogonal and non-orthogonal. Orthogonal moments produce features that are less correlated than their non-orthogonal counter parts. Further, the orthogonality property enables simple, accurate signal reconstruction from the generated moments. Moments that are non-orthogonal tend to be simpler to implement, computationally less expensive and include descriptors that have a range of useful properties i.e. scale, translation and rotation invariance. However, their highly correlated features make reconstruction more difficult.

There have been many studies using two dimensional moments for image recognition purposes. For example Dudani [17] used moments to recognise aeroplane silhouettes with results that were more successful than the human eye. Ryo [78] used local moments to recognise hand poses, while Beardsley [2] used Cartesian moments to produce simple hand control of a toy robot.

However, all of these are only interested in processing single images. Little [46] used moments to characterise optical flows between images for gait recognition. This technique still only links adjacent images, and does not consider the complete

sequence. Rosales [68] described motion by producing one image that contained information from a complete sequence, building on the work done by Davis [15]. Rosales's system was based on Hu invariant moments and was used to recognise types of motion, eg. sitting down or kicking. However, due to several images being compressed into one, subtle differences between subjects are lost due to self occlusion and overlapping of data.

For this work we started by looking at a traditional statistical method of moments to describe the motion of a person through multiple images. Unfortunately this does not provide a very detailed description of the motion, as there is no information linking the images of the sequence, since they are treated as separate entities. By using the general theory of moments a method has been developed that not only contains information about the pixel structure of the subject, but also how their movement flows *between* images.

1.5 Contributions

We will describe a novel framework for statistically characterising shapes within a temporal sequence, called velocity moments. The new framework is based around the well established centralised moments. It fuses a per-image description of a shape's structure with a description that contains information about the motion of the shape as it moves within the image sequence - exploiting temporal correlation. We propose two variations of the velocity moments, Cartesian and Zernike, each with beneficial properties. The Cartesian velocity moments are a non-orthogonal descriptor (in terms of each separate image), which are simple to compute and computationally inexpensive. However, due to their non-orthogonality the features produced are highly correlated. Discrimination within a small dataset was found to be straightforward using both structural and motion based features. When applying the technique to a large dataset, the high feature correlation may hamper the separation of different classes. The solution to this problem is to use an orthogonal basis set within the velocity moment framework - the Zernike velocity moments. The features produced by the Zernike velocity moments are invariant to image by image scale changes, making them useful in situations where camera zoom is apparent. The performance of this new framework has been analysed using synthetic data and also by application to human gait analysis. The application based properties of the moments have been analysed along with their capability to enable classification of temporal sequences.

1.6 Thesis overview

The introduction, development and characterisation of the new velocity moments, are the central aims of this thesis. This includes the background material related to

the new techniques, along with their application to human gait analysis. The thesis is separated into six chapters. Chapter 2 covers the background theory of statistical moments, as applied to images. This ranges from the theory behind moments through to implementation issues. Chapter 3 introduces the new velocity moments, the ideas behind them and their application. Preliminary analysis is included to highlight the advantages and uses of these moments in describing temporal image sequences. In Chapter 4 the velocity moments are applied to the problem of computer driven human gait classification, via a selection of different gait databases. The databases include those created in-house along with those kindly supplied by overseas research groups. Following on, Chapter 5 looks into the performance analysis of the Zernike velocity moments, via use of the human gait analysis from the previous chapter. Chapter 6 suggests possible future directions for this research. Finally Chapter 7 concludes this thesis with a discussion on the overall conclusions of this research.

1.7 Publications related to this research

There are currently four publications associated with this thesis. The early work of the Cartesian velocity moments was presented in [72], applied to synthetic images of moving shapes and a small four subject human gait database. [74] presents the analysis of a larger six subject gait database, including the analysis of optical flow images. [73] is a theory based paper describing the reconstruction of images from moments, first explaining for Cartesian moments and then extending the theory for Cartesian velocity moments. [71] describes the theory behind the Zernike velocity moments, their application to a six subject silhouette gait database and preliminary occlusion analyses.

Chapter 2

Background theory

Moments are applicable to many different aspects of image processing, ranging from invariant pattern recognition and image encoding to pose estimation. When applied to images, they describe the image content (or distribution) with respect to its axes. They are designed to capture both global and detailed geometric information about the image. Here we are using them to characterise a grey level image so as to extract properties that have analogies in statistics or mechanics. In continuous form an image can be considered as a two-dimensional Cartesian density distribution function $f(x, y)$. With this assumption, the general form of a moment of order $(p + q)$, evaluating over the complete image plane ξ is:

$$M_{pq} = \int\int_{\xi} \psi_{pq}(x, y) f(x, y)\ dx\ dy \qquad ; \qquad p, q = 0, 1, 2, ..., \infty \tag{2.1}$$

The *weighting kernel* or *basis* function is ψ_{pq}. This produces a weighted description of $f(x, y)$ over the entire plane ξ. The basis functions may have a range of useful properties that may be passed onto the moments, producing descriptions which can be invariant under rotation, scale, translation and orientation. To apply this to digital images, Equation 2.1 needs to be expressed in discrete form. The probability density function (of a continuous distribution) is different from that of the probability of a discrete distribution. For simplicity we assume that ξ is divided into square pixels of dimensions 1×1, with constant intensity I over each square pixel. So if P_{xy} is a discrete pixel value then:

$$P_{xy} = I(x, y)\Delta A \tag{2.2}$$

where ΔA is the sample or pixel area equal to one. Thus, analysing over the complete discrete image intensity plane produces:

$$M_{pq} = \sum_{x} \sum_{y} \psi_{pq}(x, y) P_{xy} \qquad ; \qquad p, q = 0, 1, 2, ..., \infty \tag{2.3}$$

The choice of basis function depends on the application and any desired invariant properties. The choice of basis may introduce constraints including limiting the x and y range, or translating the description and image to polar co-ordinates (eg. mapping it to the unit disc).

2.1 The moment generating function

To describe the distribution of a random variable the *characteristic function* can be used [61]:

$$X(w) = \int_{-\infty}^{\infty} f(x) \exp(jwx)\ dx = E[\exp(jwx)] \tag{2.4}$$

shown here for the signal density $f(x)$, where $j = \sqrt{-1}$ and w is the spatial frequency. This is essentially the Fourier transform of the signal and has a maximum at the origin $w = 0$, as $f(x) \geq 0$. Figure 2.1 shows an example of $X(w)$ for a zero mean, unit variance Gaussian density $f(x)$.

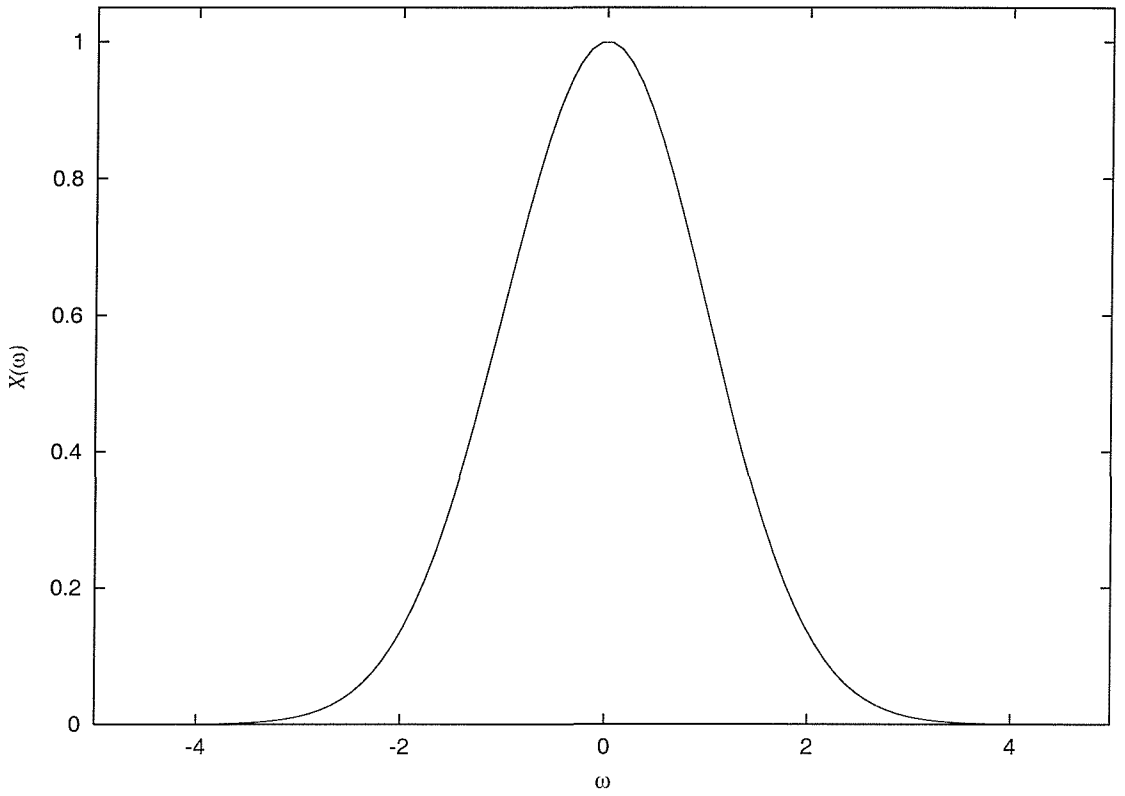


Figure 2.1: Characteristic function of a Gaussian density $f(x)$.

If $f(x)$ is the density of a positive, real valued random variable x, such that $x \epsilon \mathbb{R}$, then a continuous exponential distribution can be defined. Replacing jw in Equation 2.4 with s produces a real valued integral of the form:

$$M^x(s) = \int_{-\infty}^{\infty} f(x) \exp(xs)\ dx = E[\exp(xs)] \tag{2.5}$$

where $E[.]$ is the expectation and $M^x(s)$ exists as a real number. $M^x(s)$ is called the *moment generating function*, shown here for a one-dimensional distribution. It is used to characterise the distribution of an ergodic signal. Expressing the exponential in terms of an expanded Taylor series produces:

$$\exp(xs) = \sum_{n=0}^{\infty} \frac{x^n s^n}{n!} = 1 + xs + \frac{1}{2!}x^2 s^2 + ... + R_n(x) \tag{2.6}$$

where $R_n(x)$ is the error term. It can be seen that the series will only converge and represent $x(s)$ completely if $R_n(x) = 0$. Therefore, if the distribution is finite in length, all values outside this length must be zero (or in terms of an image, all values outside the sampled image plane must be zero). Assuming this and substituting Equation 2.6 into Equation 2.5 produces:

$$\begin{aligned} M^x(s) &= \int_{-\infty}^{\infty} f(x)\exp(xs)\ dx \\ &= \int_{-\infty}^{\infty} (1 + xs + \frac{1}{2!}x^2 s^2 + ...) f(x)\ dx \\ &= 1 + sm_1 + \frac{1}{2!}s^2 m_2 + ..., \end{aligned} \tag{2.7}$$

where m_n is the n^{th} moment about the origin. Differentiating Equation 2.5 n times with respect to s produces:

$$M_n^x(s) = E[x^n \exp(xs)] \tag{2.8}$$

If $M^x(s)$ is differentiable at zero, then the n^{th} order moments about the origin are given by:

$$M_n^x(0) = E[x^n] = m_n \tag{2.9}$$

So the first three moments of this distribution are:

$$\begin{aligned} M_0^x(s) &= E[\exp(xs)] \ ; \ M_0^x(0) = 1 \\ M_1^x(s) &= E[x\exp(xs)] \ ; \ M_1^x(0) = x \\ M_2^x(s) &= E[x^2\exp(xs)] \ ; \ M_2^x(0) = x^2 \end{aligned} \tag{2.10}$$

If the distribution of the signal is a Gaussian, then it is completely described by its two moments, mean $(M_1^x(0))$ and variance $(M_2^x(0) - (M_1^x(0))^2)$, while the total area $(M_0^x(0))$ is 1. If the joint moment $M^{xy}(s)$ for two signals is required (i.e. a two-dimensional image) then it is noted that:

$$M^{xy}(s) = E[\exp((x+y)s)] = E[\exp(xs)\exp(ys)] \tag{2.11}$$

and assuming that x and y are independent, then:

$$M^{xy}(s) = E[\exp(xs)]E[\exp(ys)] = M^{x}(s)M^{y}(s) \tag{2.12}$$

In conclusion, it is possible to evaluate the moments of a distribution by two methods. Either by using the direct integration (Equation 2.1), or by use of the moment generating function (Equation 2.5). However, in practice the moment generating function is more widely applied to the problem of calculating moment invariants, while the direct integration method is used to calculate specific moment values.

2.2 Non-orthogonal moments

Hu [27, 28], stated that the continuous two-dimensional $(p+q)^{th}$ order Cartesian moment is defined in terms of Riemann integrals as:

$$m_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^{p}y^{q}\ f(x,y)\ dx\ dy \tag{2.13}$$

It is assumed that $f(x,y)$ is a piecewise continuous, bounded function and that it can have non-zero values only in the finite region of the $x-y$ plane (i.e. all values outside the image plane are zero - see the Taylor series expansion (Equation 2.6) and explanation in the previous section). If this is so, then moments of all orders exist and the following uniqueness theorem holds [28]:

Theorem 1 *Uniqueness theorem : the moment sequence m_{pq} (Equation 2.13 - the basis $x^{p}y^{q}$) is uniquely defined by $f(x,y)$ and conversely, $f(x,y)$ is uniquely defined by m_{pq}.*

This implies that the original image can be described and reconstructed, if sufficiently high order moments are used. By adapting Equation 2.5 to two dimensions, the Cartesian moments (Equation 2.13) can be expressed in terms of the moment generating function. Analysing a two-dimensional irradiance distribution $f(x,y)$:

$$M^{xy}(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(ux+vy)f(x,y)\ dx\ dy \tag{2.14}$$

and expanding the exponential using Taylor series produces:

$$M^{xy}(u,v) = \sum_{p=0}^{\infty}\sum_{q=0}^{\infty} \frac{u^{p}}{p!}\frac{v^{q}}{q!}m_{pq} \tag{2.15}$$

where m_{pq} are the moments of this two dimensional distribution.

2.2.1 *Cartesian moments*

The discrete version of the Cartesian moment (Equation 2.13) for an image consisting of pixels P_{xy}, replacing the integrals with summations, is:

$$m_{pq} = \sum_{x=1}^{M} \sum_{y=1}^{N} x^p y^q P_{xy} \tag{2.16}$$

Where M and N are the image dimensions and the monomial product $x^p y^q$ is the basis function. Figure 2.2 illustrates the non-orthogonal (highly correlated) nature of these monomials (in contrast to the orthogonal polynomials in Figure 2.6, to be discussed later) plotted for the positive x axis only. The zero order moment m_{00}

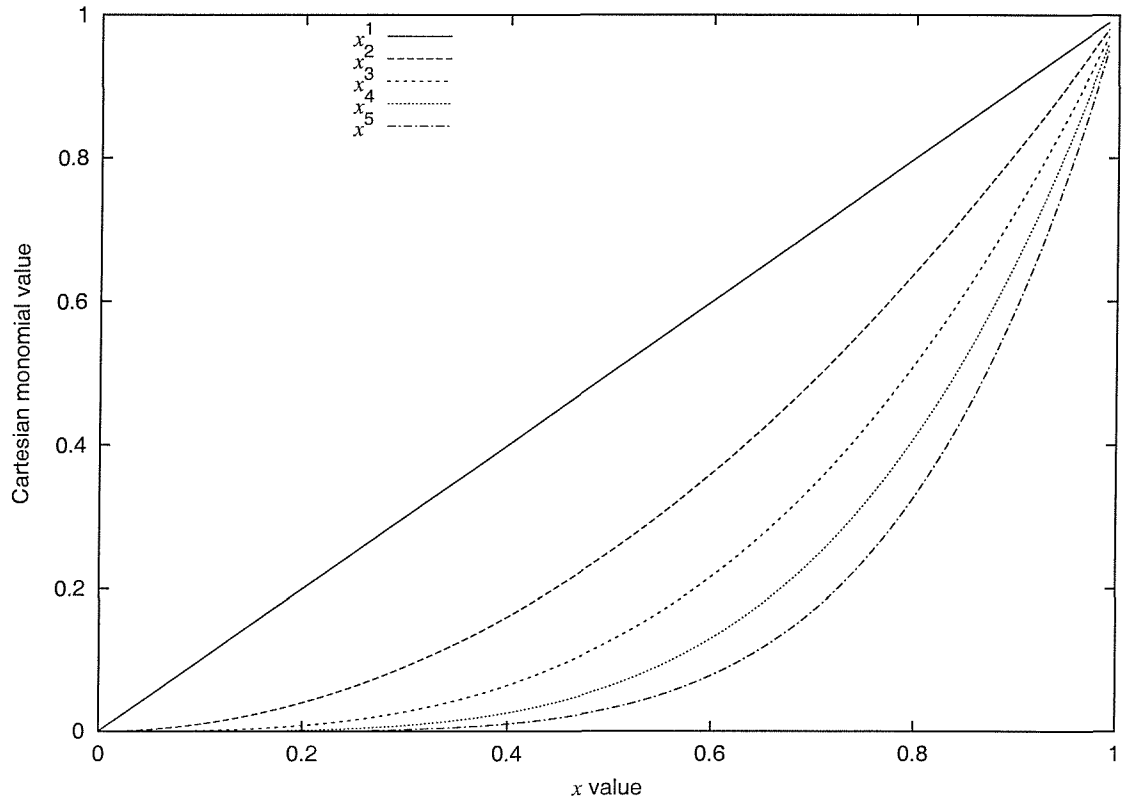


Figure 2.2: The first five Cartesian monomials.

is defined as the total mass (or power) of the image. If this is applied to a binary (i.e. a silhouette) $M \times N$ image of an object, then this is literally a pixel count of the number of pixels comprising the object.

$$m_{00} = \sum_{x=1}^{M} \sum_{y=1}^{N} P_{xy} \tag{2.17}$$

The two first order moments are used to find the Centre Of Mass (COM) of an image. If this is applied to a binary image and the results are then normalised with respect to the total mass (m_{00}), then the result is the centre co-ordinates of the object. Accordingly, the centre co-ordinates $\overline{x}, \overline{y}$ are given by :

$$\overline{x} = \frac{m_{10}}{m_{00}} \qquad \overline{y} = \frac{m_{01}}{m_{00}} \tag{2.18}$$

The COM describes a unique position within the field of view which can then be used to compute the centralised moments of an image.

2.2.2 Centralised moments

The definition of a discrete centralised moment as described by Hu[28] is:

$$\mu_{pq} = \sum_{x=1}^{M}\sum_{y=1}^{N}(x-\overline{x})^p(y-\overline{y})^q P_{xy} \tag{2.19}$$

This is essentially a translated Cartesian moment, which means that the centralised moments are invariant under translation. To enable invariance to scale, normalised moments η_{pq} are used [83], given by:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}} \tag{2.20}$$

where :

$$\gamma = \frac{p+q}{2} + 1 \qquad \forall(p+q) \geq 2 \tag{2.21}$$

2.2.3 Image reconstruction

Having described an image by a set of moments, it may prove useful to investigate which moments give rise to which characteristics of the image, or vice versa. This can be achieved by reconstructing the original image from the calculated moments. Moment reconstruction, for moments with orthogonal basis functions (such as Legendre and Zernike moments which will be discussed later) has been developed extensively, [62, 65, 79, 80]. However, where the basis set is non-orthogonal (such as Cartesian and centralised moments), only one method has appeared (although, non-orthogonal transform methods exist). This is the method of moment matching for non-orthogonal moment reconstruction [79]. The method is based upon creating a continuous function that has identical moments to that of the original function. In this section it has been applied first to Cartesian moments and then to the centralised moments. It must be noted that in applying the theory to sampled images, the continuous conditions are replaced by discrete versions, reducing the accuracy of the final function.

Assuming that all moments M_{pq} of a function $f(x,y)$ and of order $N=(p+q)$ are known from zero through to order N_{max}, it is then possible to obtain the continuous function $g(x,y)$ whose moments match those of the original function $f(x,y)$, up to order N_{max}. (With reference to the Taylor series expansion in Section 2.1), assuming that the given continuous function can be defined as:

$$g(x,y) = g_{00} + g_{10}x + g_{01}y + g_{20}x^2 + g_{11}xy + \ ...g_{pq}x^p y^q \tag{2.22}$$

which reduces to:

$$g(x,y) = \sum_{p=0}\sum_{q=0} g_{pq}x^p y^q \quad ; \quad N_{max} = p+q \tag{2.23}$$

then the constant coefficients g_{pq}, are calculated, so that the moments of $g(x,y)$ match those of $f(x,y)$, assuming that the image is a continuous function bounded by:

$$x \ \epsilon \ [-1,1] \ , \ \ y \ \epsilon \ [-1,1] \tag{2.24}$$

These limits can be achieved by normalising the pixel range over which the Cartesian moments are calculated, thus:

$$\int_{-1}^{1}\int_{-1}^{1} g(x,y)x^p y^q \ dx \ dy \equiv M_{pq} \tag{2.25}$$

Substituting Equation 2.22 into Equation 2.25 and then solving the integration produces a set of Linear Equations (LE), the number of which is determined by the order $(p+q)$ of reconstruction. These can then be solved for the coefficients g_{pq} (in terms of the moments M_{pq}) by using matrix inversion. For order three ($(p+q) \leq 3$), the LEs in matrix form are:

$$\begin{bmatrix} 1 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{5} & \frac{1}{9} \\ \frac{1}{3} & \frac{1}{9} & \frac{1}{5} \end{bmatrix} \begin{bmatrix} g_{00} \\ g_{20} \\ g_{02} \end{bmatrix} = \frac{1}{4} \begin{bmatrix} M_{00} \\ M_{20} \\ M_{02} \end{bmatrix} \tag{2.26}$$

$$\begin{bmatrix} \frac{1}{3} & \frac{1}{5} & \frac{1}{9} \\ \frac{1}{5} & \frac{1}{7} & \frac{1}{15} \\ \frac{1}{9} & \frac{1}{15} & \frac{1}{15} \end{bmatrix} \begin{bmatrix} g_{10} \\ g_{30} \\ g_{12} \end{bmatrix} = \frac{1}{4} \begin{bmatrix} M_{10} \\ M_{30} \\ M_{12} \end{bmatrix} \tag{2.27}$$

$$\begin{bmatrix} \frac{1}{3} & \frac{1}{5} & \frac{1}{9} \\ \frac{1}{5} & \frac{1}{7} & \frac{1}{15} \\ \frac{1}{9} & \frac{1}{15} & \frac{1}{15} \end{bmatrix} \begin{bmatrix} g_{01} \\ g_{03} \\ g_{21} \end{bmatrix} = \frac{1}{4} \begin{bmatrix} M_{01} \\ M_{03} \\ M_{21} \end{bmatrix} \tag{2.28}$$

and finally:

$$g_{11} = \frac{9}{4} M_{11} \tag{2.29}$$

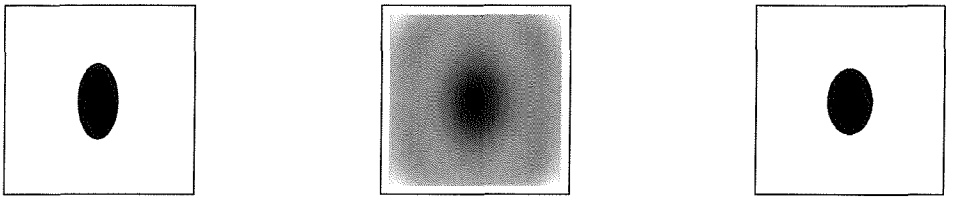
Applying matrix inversion to the first matrix, Equation 2.26 produces:

$$\begin{bmatrix} 14 & -15 & -15 \\ -15 & 45 & 0 \\ -15 & 0 & 45 \end{bmatrix} \begin{bmatrix} M_{00} \\ M_{20} \\ M_{02} \end{bmatrix} = \frac{1}{16} \begin{bmatrix} g_{00} \\ g_{20} \\ g_{02} \end{bmatrix} \tag{2.30}$$

By repeating this for all the matrices, it is possible to calculate all the coefficients. If they are then substituted back into Equation 2.22 an expression for $g(x, y)$ is produced. This expression can then used to reconstruct an approximation of the original image. The reconstruction function $g(x, y)$ is now in terms of weighted sums of the moments M_{pq}, which have been previously calculated from the original function $f(x, y)$. The resultant function $g(x, y)$ for order three is:

$$\begin{aligned} 16g(x,y) &= (14M_{00} - 15M_{20} - 15M_{02}) \\ &+ (90M_{10} - 105m_{30} - 45M_{12})x \\ &+ (90M_{01} - 105M_{03} - 45M_{21})y \\ &+ (-15M_{00} + 45M_{20})x^2 + 36M_{11}xy \\ &+ (-15M_{00} + 45M_{02})y^2 + (-105M_{10} + 175M_{30})x^3 \\ &+ (-45M_{01} + 135M_{21})x^2y + (-45M_{10} + 135M_{12})xy^2 \\ &+ (-105M_{01} + 175M_{03})y^3 \end{aligned} \tag{2.31}$$

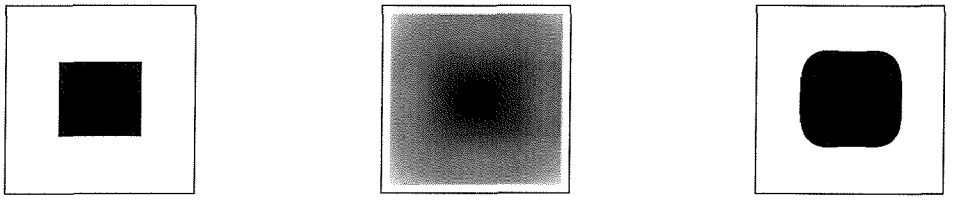
Implementing this method to order $(p+q) = 8$ for binary images of simple shapes produces recognisable results, as shown in Figures 2.3 and 2.4. Figure 2.3a is the original image from which the moments were calculated and Figure 2.3b is the image reconstructed from the moments. The borders of the shape appear unclear, but they appear when the reconstructed image is thresholded, Figure 2.3c. Here the level of the applied threshold was adjusted by visual comparison with the original image. Due to the nature of the continuous function, the final shape is dependent on the threshold level, as is apparent in Figure 2.3c, as compared with the original image, Figure 2.3a. This analysis is then repeated for the rectangle in Figure 2.4. The corners of the rectangle in Figure 2.4c are missing. The corners represent the high frequency content in the image, thus will be described fully by higher order moments. So the thresholded shape will converge to the original shape as the number of moments (and thus the order) increases. However for more complex shapes, higher accuracy $(p + q) \gg 8$ is needed. This is analogous to the high frequency information needed to reconstruct pulsed time domain waveforms, using



(a) Original Image (b) Reconstructed continuous function (c) Thresholded

Figure 2.3: Order 8 Cartesian reconstruction of an ellipse.

methods like Fourier series. As the order (and accuracy) increases, so does the number of LEs that need to be solved (reconstruction for order eight resulted in forty five LE's). Further, if it was required to increase the order of reconstruction (using Equation 2.31), then all coefficients need to be re-calculated. This is due to the correlated nature of the Cartesian moments, each moment does not simply provide its own individual contribution, (unlike the orthogonal case which will be discussed later in this chapter). It is interesting to note the effects of the Gibbs phenomena [75] which are more evident in the reconstructed ellipse - Figure 2.3b. The Gibbs phenomena (explained in terms of Fourier series) is the inability for a continuous function to recreate a step function - no matter how many finite high order terms are used, an overshoot of the function will occur. Here the discontinuous edge of the original intensity function of the ellipse (between the ellipse and the background) appear unclear in the reconstruction. While outside of the original area of the ellipse, 'ripples' of overshoot of the continuous function are visible.



(a) Original Image (b) Reconstructed continuous function (c) Thresholded

Figure 2.4: Order 8 Cartesian reconstruction of a rectangle.

By assuming the same constraints as for Cartesian moment matching, the theory can be extended to centralised moments. The continuous function $g(x, y)$ is now defined as:

$$g(x,y) = \sum_{p=0}\sum_{q=0} g_{pq}(x-\overline{x})^p(y-\overline{y})^q \quad ; \quad N_{max} = p+q \tag{2.32}$$

similarly, Equation 2.25 becomes:

$$\int_{-1}^{1}\int_{-1}^{1} g(x,y)(x-\overline{x})^p(y-\overline{y})^q\,dx\,dy \equiv M_{pq} \tag{2.33}$$

where $\overline{x}$ and $\overline{y}$ are the x and y COM's, respectively. Solving for $g(x,y)$ is then achieved in the same manner as already described for the Cartesian case.

2.2.4 Symmetry properties

A measure of asymmetry in an image is given by its *skewness*, where here the description is a statistical measure of a distribution's degree of deviation from symmetry about the mean [53]. The third order moments (skewness and bi-correlations) will be zero if the distribution is symmetric eg. Gaussian. The degree of skewness can be determined using the two third order moments, μ_{30} and μ_{03}. Prokop [65] used these moments as a basis to define the coefficients of skewness. The direction of skewness can be determined by analysing the signs of these results.

More generally, Li [43] described the basis function $x^p y^q$ (in Equation 2.16), as a weighting function which extracts features of the image $f(x,y)$ concerning the symmetry in the irradiance distribution. Li used this property to show how low order $(p+q)^{th}$ normalised centralised moments (Equation 2.20) produce descriptions which are directly comparable to the existence of symmetry within the image. Here symmetry is being detected about the COM of the image, hence the use of the centralised moments. The first seven scale-normalised centralised moments $(\eta_{11}, \eta_{20}, \eta_{02}, \eta_{21}, \eta_{12}, \eta_{30}, \eta_{03})$ were analysed using typed characters as binary input images. It was shown that by looking at the sign and the magnitude of the centralised moments, character recognition based on symmetry properties is possible. Here follows a summary of this work. Shapes that are either symmetric about the x or y axes will produce $\eta_{11} = 0$. For shapes symmetrical about the y axis $\eta_{12} = 0$ and $\eta_{30} = 0$, Figure 2.5a and Table 2.1. However for shapes symmetric about the x axis, $\eta_{03} = 0$ and η_{12} is positive, Figure 2.5b and Table 2.1. Further to this the following generalities are true:

$$\eta_{pq} = 0 \qquad \forall p = 0,2,4.. \;\; ; \;\; q = 1,3,5.. \tag{2.34}$$

for shapes symmetric about the x axis. However shapes which are asymmetrical about the x axis produce:

$$\eta_{pq} < 0 \;\; \forall p = 0,2,4.. \;\; ; \;\; q = 1,3,5.. \tag{2.35}$$

and:

$$\eta_{p0} > 0 \;,\; \eta_{0p} > 0 \qquad \forall p = 0,2,4.. \quad f(x,y) > 0 \tag{2.36}$$

In this way it can seen that the sign of the normalised centralised moments can be arranged to give *qualitative* information about the shape being described (i.e. the existence of symmetry), while the magnitude of the centralised moments gives a *quantitative* description (i.e. their size and density).

(a) (b)

Figure 2.5: Axes of symmetry for typed characters.

Character	η_{11}	η_{20}	η_{02}	η_{21}	η_{12}	η_{30}	η_{03}
M	0	+	+	-	0	0	-
C	0	+	+	0	+	+	0

Table 2.1: Typed characters η_{pq} values indicating symmetry.

2.3 Hu invariant set

The non-orthogonal centralised moments are translation invariant and can be normalised with respect to changes in scale. However, to enable invariance to rotation they require reformulation. Hu [28] described two different methods for producing rotation invariant moments. The first used a method called principal axes, however it was noted that this method can break down when images do not have unique principal axes. Such images are described as being rotationally symmetric. The second method Hu described is the method of absolute moment invariants and is discussed here. Hu derived these expressions from algebraic invariants applied to the moment generating function under a rotation transformation. They consist of groups of nonlinear centralised moment expressions. The result is a set of absolute orthogonal (i.e. rotation) moment invariants, which can be used for scale, position, and rotation invariant pattern identification. These were used in a simple pattern recognition experiment to successfully identify various typed characters. They are computed from normalised centralised moments up to order three and are shown below:

$$I_1 = \eta_{20} + \eta_{02} \tag{2.37}$$

$$I_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \tag{2.38}$$

$$I_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \tag{2.39}$$

$$I_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \tag{2.40}$$

$$I_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (2.41)$$

$$I_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})] \quad (2.42)$$

Finally a skew invariant, to help distinguish mirror images, is:

$$I_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (2.43)$$

These moments are of finite order, therefore, unlike the centralised moments they do not comprise a complete set of image descriptors, [43]. However, higher order invariants can be derived, [3, 28]. It should be noted that this method also breaks down, as with the method based on the principal axis for images which are rotationally symmetric as the seven invariant moments will be zero [65].

2.4 Orthogonal moments

Cartesian moments, Equation 2.13 are formed using a monomial basis set $x^p y^q$. This basis set is non-orthogonal and this property is passed onto the Cartesian moments. These monomials increase rapidly in range as the order increases, producing highly correlated descriptions. This can result in important descriptive information being contained within small differences between moments, which leads to the need for high computational precision.

However, moments produced using orthogonal basis sets exist. These orthogonal moments have the advantage of needing lower precision to represent differences to the same accuracy as the monomials. The orthogonality condition simplifies the reconstruction of the original function from the generated moments. Orthogonality means mutually perpendicular, expressed mathematically - two functions y_m and y_n are orthogonal over an interval $a \leq x \leq b$ if and only if:

$$\int_a^b y_m(x)\ y_n(x)dx = 0 \quad ; \quad m \neq n \quad (2.44)$$

Here we are primarily interested in discrete images, so the integrals within the moment descriptors are replaced by summations. It is noted that a sequence of polynomials which are orthogonal with respect to integration, are also orthogonal with respect to summation, [85]. Two such (well established) orthogonal moments are Legendre and Zernike.

2.4.1 Legendre moments

The Legendre moments [79] of order $(m+n)$ are defined as:

$$\lambda_{mn} = \frac{(2m+1)(2n+1)}{4} \int_{-1}^{1} \int_{-1}^{1} P_m(x) P_n(y) f(x,y)\, dx\, dy \tag{2.45}$$

where $m, n = 0, 1, 2, ..., \infty$, P_m and P_n are the Legendre polynomials and $f(x,y)$ is the continuous image function. The Legendre polynomials are a complete orthogonal basis set defined over the interval $[-1, 1]$. For orthogonality to exist in the moments, the image function $f(x,y)$ is defined over the same interval as the basis set, where the n^{th} order Legendre polynomial is defined as:

$$P_n(x) = \sum_{j=0}^{n} a_{nj}\, x^j \tag{2.46}$$

and a_{nj} are the Legendre coefficients given by:

$$a_{nj} = (-1)^{(n-j)/2} \frac{1}{2^n} \frac{(n+j)!}{(\frac{(n-j)}{2})!(\frac{(n+j)}{2})!j!} \qquad \text{where } n-j = \text{even} \tag{2.47}$$

So, for a discrete image with current pixel P_{xy}, Equation 2.45 becomes:

$$\lambda_{mn} = \frac{(2m+1)(2n+1)}{4} \sum_{x} \sum_{y} P_m(x) P_n(y) P_{xy} \tag{2.48}$$

and x, y are defined over the interval $[-1, 1]$.

2.4.2 Complex Zernike moments

The Zernike polynomials were first proposed in 1934 by Zernike [87]. Their moment formulation appears to be one of the most popular, outperforming the alternatives [80] (in terms of noise resilience, information redundancy and reconstruction capability). The pseudo-Zernike formulation proposed by Bhatia and Wolf [6] further improved these characteristics. However, here we study the original formulation of these orthogonal invariant moments.

Complex Zernike moments [79] are constructed using a set of complex polynomials which form a complete orthogonal basis set defined on the unit disc $(x^2+y^2) \leq 1$. They are expressed as:

$$A_{mn} = \frac{m+1}{\pi} \int_x \int_y f(x,y)[V_{mn}(x,y)]^*\, dx\, dy \qquad \text{where } x^2+y^2 \leq 1 \tag{2.49}$$

where $m = 0, 1, 2, ..., \infty$ and defines the order, $f(x,y)$ is the function being described and $*$ denotes the complex conjugate. While n is an integer (that can be positive or

negative) depicting the angular dependence, or rotation, subject to the conditions:

$$m - |\,n\,| \;=\; even \;\;,\;\;\; |\,n\,| \;\leq\; m \tag{2.50}$$

and $A^*_{mn} = A_{m,-n}$ is true. The Zernike polynomials [87] $V_{mn}(x, y)$ expressed in polar coordinates are:

$$V_{mn}(r, \theta) \;=\; R_{mn}(r) \exp(jn\theta) \tag{2.51}$$

where (r, θ) are defined over the unit disc, $j = \sqrt{-1}$ and $R_{mn}(r)$ is the orthogonal radial polynomial, defined as:

$$R_{mn}(r) = \sum_{s=0}^{\frac{m-|n|}{2}} (-1)^s \;\; F(m, n, s, r) \tag{2.52}$$

where:

$$F(m, n, s, r) = \frac{(m-s)!}{s!(\frac{m+|n|}{2} - s)!\;(\frac{m-|n|}{2} - s)!}\; r^{m-2s} \tag{2.53}$$

where $R_{mn}(r) = R_{m,-n}(r)$ and it must be noted that if the conditions in Equation 2.50 are not met, then $R_{mn}(r) = 0$. The first six orthogonal radial polynomials are:

$$\begin{aligned} R_{00}(r) &= 1 & R_{11}(r) &= r \\ R_{20}(r) &= 2r^2 - 1 & R_{22}(r) &= r^2 \\ R_{31}(r) &= 3r^3 - 2r & R_{33}(r) &= r^3 \end{aligned} \tag{2.54}$$

Figure 2.6 shows eight such radial responses, where it can been seen that the polynomials become more grouped, as they approach the edge of the unit disc (r approaches unity). (Care must be taken with regard to the accuracy of these polynomial calculations as the factorial operations can quickly produce large integer values, even at relatively low order m). The difference between these orthogonal polynomials and the non-orthogonal monomials can be seen by comparing Figure 2.6 with Figure 2.2. So for a discrete image, if P_{xy} is the current pixel then Equation 2.49 becomes:

$$A_{mn} = \frac{m+1}{\pi} \sum_x \sum_y P_{xy}[V_{mn}(x, y)]^* \qquad \text{where } x^2 + y^2 \leq 1 \tag{2.55}$$

To calculate the Zernike moments, the image (or region of interest) is first mapped to the unit disc using polar coordinates, where the centre of the image is the origin of the unit disc. Those pixels falling outside the unit disc are not used in the calculation. The coordinates are then described by the length of the vector from the origin to the coordinate point, r, and the angle from the x axis to the vector r,

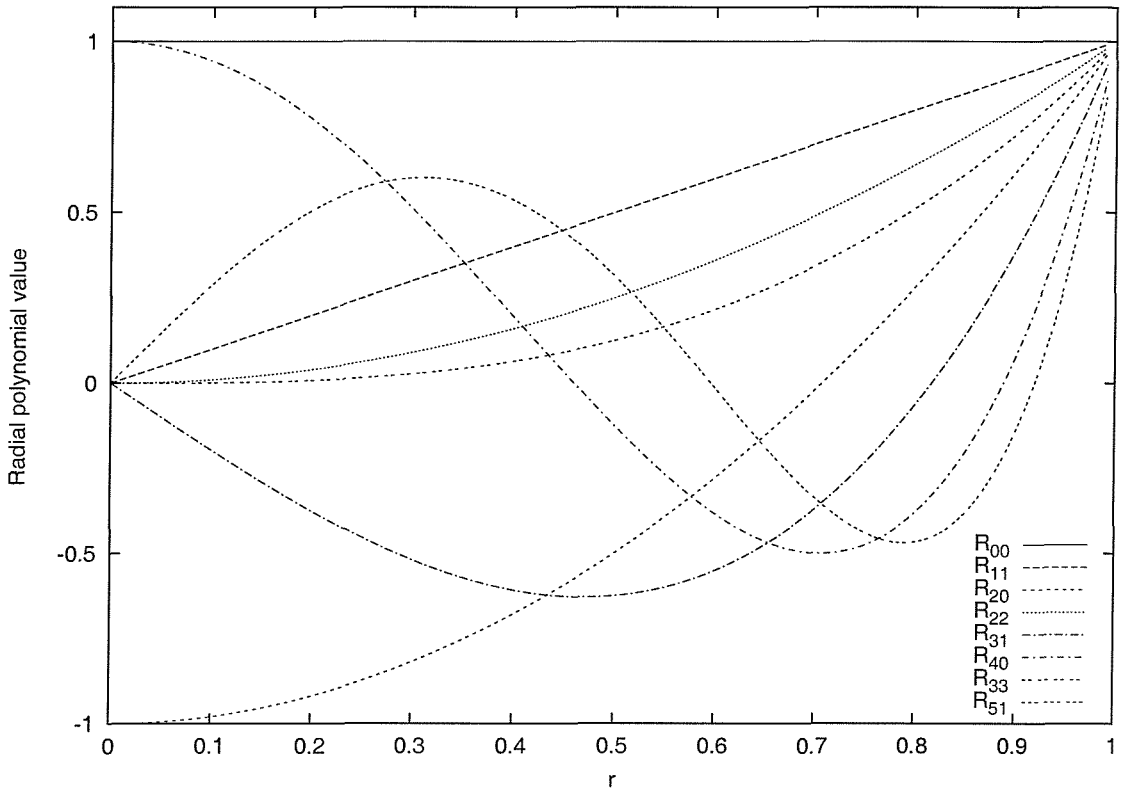


Figure 2.6: Eight orthogonal radial polynomials plotted for increasing r.

θ, by convention measured from the positive x axis in a counter clockwise direction. The mapping from Cartesian to polar coordinates is:

$$x = r\ \cos\theta \qquad y = r\ \sin\theta \tag{2.56}$$

where

$$r = \sqrt{x^2 + y^2} \qquad \theta = \tan^{-1}\left(\frac{y}{x}\right) \tag{2.57}$$

However, $\tan^{-1}$ in practice is often defined over the interval $-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$, so care must be taken as to which quadrant the Cartesian coordinates appear in. Translation and scale invariance can be achieved by normalising the image using the Cartesian moments prior to calculation of the Zernike moments [37]. Translation invariance is achieved by moving the origin to the image's COM, causing $m_{01} = m_{10} = 0$. Following this, scale invariance is produced by altering each object so that its area (or pixel count for a binary image) is $m_{00} = \beta$, where β is a predetermined value. Both invariance properties (for a binary image) can be achieved using :

$$h(x, y) = f\left(\frac{x}{a} + \overline{x}, \frac{y}{a} + \overline{y}\right) \qquad \text{where } a = \sqrt{\frac{\beta}{m_{00}}} \tag{2.58}$$

and $h(x, y)$ is the new translated and scaled function. The error involved in the

discrete implementation can be reduced by interpolation. If the coordinate calculated by Equation 2.58 does not coincide with an actual grid location, the pixel value associated with it is interpolated from the four surrounding pixels. As a result of the normalisation, the Zernike moments $\mid A_{00} \mid$ and $\mid A_{11} \mid$ are set to known values. $\mid A_{11} \mid$ is set to zero, due to the translation of the shape to the centre of the coordinate system. This however will be affected by a discrete implementation where the error in the mapping will decrease as the shape (being mapped) size (or pixel-resolution) increases. $\mid A_{00} \mid$ is dependent on m_{00}, and thus on β:

$$\mid A_{00} \mid \; = \frac{\beta}{\pi} \tag{2.59}$$

Further, the absolute value of a Zernike moment is rotation invariant as reflected in the mapping of the image to the unit disc. The rotation of the shape around the unit disc is expressed as a phase change, if ϕ is the angle of rotation, A^R_{mn} is the Zernike moment of the rotated image and A_{mn} is the Zernike moment of the original image then:

$$A^R_{mn} = A_{mn} \exp(-jn\phi) \tag{2.60}$$

A note on image encoding

By returning to the continuous form of the Zernike moments (in order to enable easy manipulation of the moments), it is possible to gain an insight into the image encoding itself. As before, the Zernike moments are defined by Equation 2.49, where the Zernike polynomials are expressed by Equation 2.51. Converting the integration of Equation 2.49 to polar coordinates, using $dx\ dy = r\ dr\ d\theta$ and Equations 2.51, 2.56 and 2.57, produces:

$$\begin{aligned} A_{mn} &= \frac{m+1}{\pi} \int_x \int_y f(x,y)[V_{mn}(x,y)]^*\ dx\ dy \qquad \text{where } x^2 + y^2 \leq 1 \\ &= \frac{m+1}{\pi} \int_0^{2\pi} \int_0^1 f(r,\theta) R_{mn}(r) \exp(-jn\theta)\ r\ dr\ d\theta \end{aligned} \tag{2.61}$$

Re-arranging the integral produces:

$$A_{mn} = \frac{m+1}{\pi} \int_0^{2\pi} \exp(-jn\theta) \int_0^1 f(r,\theta)\ r\ R_{mn}(r)\ dr\ d\theta \tag{2.62}$$

It can be seen that the $f(r,\theta)$ term is weighted by $rR_{mn}(r)$. The radial polynomials are normalised such that:

$$R_{mn}(1) = 1 \tag{2.63}$$

is true [6, 53]. Due to the unit disc, r is at a maximum when equal to unity, meaning that:

$$| R_{mn}(r) | \leq 1 \tag{2.64}$$

This result is illustrated in Figure 2.6. Using these results and returning to the weighted term (in Equation 2.62), we can see that the weighting is bounded by r:

$$| rR_{mn}(r) | \leq r \tag{2.65}$$

This produces descriptions which are weighted in favour of their distance from the origin of the unit disc. Those pixels lying closer to the perimeter of the unit disc will have more weight than those lying closer to the origin. As r approaches unity the radial polynomials display steeper gradients and converge (becoming more correlated - Figures 2.6 and 2.7). The higher order polynomials (and their corresponding moments) will have improved capability in describing image detail due to their increased oscillations (see Figure 2.7) especially in the region before convergence due to the increased frequency of these oscillations. Image detail which is encoded around the region of convergence will be more correlated. However, these characteristics can be exploited. Considering the case of applied perimeter noise on simple shapes (i.e. circle, square and triangle), the effects of the noise (close to the disc's perimeter) can be reduced by scaling the shape (during mapping) to an appropriate size, where the noise cannot be described efficiently enough to pose a problem, however enabling sufficient detail to describe the shape.

2.4.3 Image reconstruction

The method of moment matching (Section 2.2.3), as described for the reconstruction of non-orthogonal moments is also applicable to reconstruction of an image by orthogonal moments. However, the orthogonality condition enables a faster, more direct approach. Teague [79] showed that, for orthogonal Legendre moments, if all moments of a Cartesian function $f(x, y)$ up to a given order N_{max} are known, then it is possible to reconstruct a discrete function $\hat{f}(x, y)$, whose moments match those of the original function $f(x, y)$, up to the order N_{max}. This relationship is due to the orthogonality condition of the Legendre moments, while the accuracy of the reconstructed function improves as N_{max} approaches infinity. Khotanzad [37] expressed this relationship in terms of Zernike moments, shown here in radial coordinates:

$$\hat{f}(r, \theta) = \sum_{m=0}^{N_{max}} \sum_{n} A_{mn} V_{mn}(r, \theta) \tag{2.66}$$

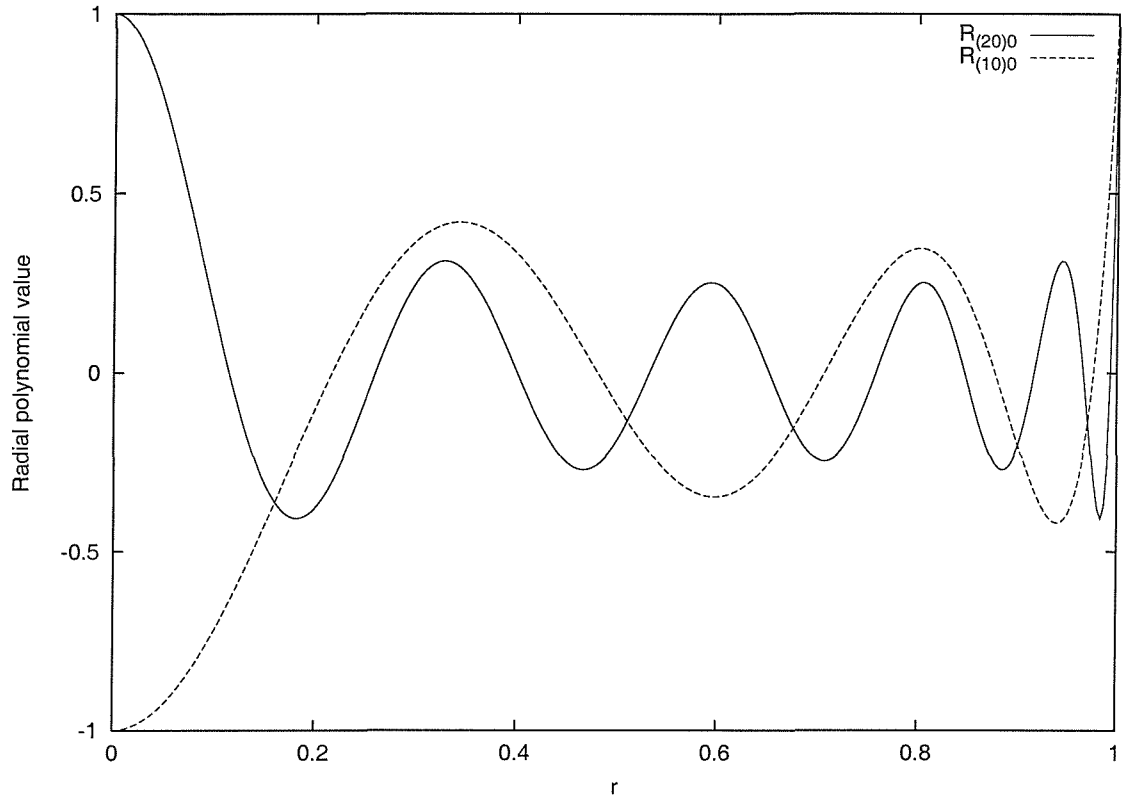


Figure 2.7: Higher order orthogonal radial polynomial plotted for increasing r.

and n is constrained by Equation 2.50. Expanding this using real-valued functions produces:

$$\hat{f}(r,\theta) = \sum_{m=0}^{N_{max}} \sum_{n>0} (C_{mn} \cos n\theta + S_{mn} \sin n\theta)\ R_{mn}(r) + \frac{C_{m0}}{2} R_{m0}(r) \tag{2.67}$$

composed of their real ($Re[.]$) and imaginary ($Im[.]$) parts:

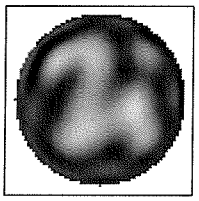
$$C_{mn} = 2Re\,[A_{mn}] = \frac{2m+2}{\pi} \sum_x \sum_y f(r,\theta) R_{mn}(r) \cos n\theta \tag{2.68}$$

$$S_{mn} = -2Im\,[A_{mn}] = \frac{-2m-2}{\pi} \sum_x \sum_y f(r,\theta) R_{mn}(r) \sin n\theta \tag{2.69}$$

bounded by $x^2 + y^2 \leq 1$. Here, each Zernike moment simply adds its own contribution to the function $\hat{f}(r,\theta)$, unlike the Cartesian reconstruction case discussed in Section 2.2.3. Figure 2.8b shows the result of order 2 through 12 reconstruction on a 64×64 image, while Figure 2.8a is the original image. Orders 0 and 1 are discarded due to the scale and translation mapping used, Equations 2.58 and 2.59. This makes $\mid A_{11} \mid$ zero, while $\mid A_{00} \mid$ (the shape's area) is set to a known value, β. Due to the nature of the function $\hat{f}(r,\theta)$, the Gibbs phenomena [75] will affect the

final result (as already mentioned for the moment matching case in Section 2.2.3). However, the effects are less apparent in Figure 2.8b due to faster convergence of the final function $\hat{f}(r, \theta)$.

(a) The original image.



(b) Order $2 - 12$ reconstruction

Figure 2.8: Zernike moment reconstruction example.

2.5 Relating Zernike and Cartesian moments

To help reduce computation complexity, it may prove useful to express the Zernike moments in terms of Cartesian moments. This removes the need for the polar mapping of the image, while also removing the dependence on the trigonometric functions. Alternatively, expressing Cartesian moments in this way would aid the selection of less correlated descriptors. This conversion can be achieved by slightly re-arranging the Zernike moment equation. If, as before, the Zernike polynomials are given by Equation 2.51 [87] and the radial polynomials $R_{mn}(r)$ are defined by Equation 2.52, re-arranging $F(m, n, s, r)$ gives:

$$\begin{aligned} F(m,n,s,r) &= \frac{(m-s)!}{s!(\frac{m+|n|}{2}-s)!\,(\frac{m-|n|}{2}-s)!}\, r^{m-2s} \\ &= \frac{(m-s)!}{s!(\frac{m-2s+|n|}{2})!\,(\frac{m-2s-|n|}{2})!}\, r^{m-2s} \end{aligned} \tag{2.70}$$

then substituting $k = (m - 2s)$ and re-arranging again, produces:

$$R_{mn}(r) = \sum_{k=n}^{m} B_{mnk}\, r^k \qquad (m-k)\ is\ even, n \geq 0\ , \tag{2.71}$$

where:

$$B_{mnk} = \frac{(-1)^{(m-k)/2}\,(\frac{m+k}{2})!}{(\frac{m-k}{2})!\,(\frac{k+n}{2})!\,(\frac{k-n}{2})!} \tag{2.72}$$

Using this manipulated form of the radial polynomials produces Zernike moment definitions (in continuous form) of:

$$A_{mn} = \frac{m+1}{\pi} \sum_{k=n}^{m} B_{mnk} \int_0^{2\pi} \int_0^1 r^k \exp(-jn\theta) f(r,\theta)\ r\ dr\ d\theta \quad ; \quad r \leq 1 \tag{2.73}$$

which when translated to Cartesian coordinates is:

$$A_{mn} = \frac{m+1}{\pi} \sum_{k=n}^{m} B_{mnk} \int_x \int_y (x - jy)^n (x^2 + y^2)^{(k-n)/2} f(x,y)\ dx\ dy \tag{2.74}$$

bounded by $x^2 + y^2 \leq 1$ and $j = \sqrt{-1}$. The double integral can now be expressed in terms of a series of summed Cartesian moments, of the form:

$$m_{pq} = \int_x \int_y x^p y^q f(x,y)\ dx\ dy \tag{2.75}$$

For example:

$$\begin{aligned} Z_{00} &= \frac{1}{\pi} \sum_{k=0}^{0} B_{00k} \int_x \int_y (x - jy)^0 (x^2 + y^2)^{k/2} f(x,y)\ dx\ dy \\ &= \frac{1}{\pi} \int_x \int_y f(x,y)\ dx\ dy \\ &= \frac{1}{\pi} m_{00} \end{aligned} \tag{2.76}$$

It must be noted that this comparison is only valid if the Cartesian moments are calculated on images confined to $[-1, 1]$, which is due to the Zernike moments being calculated over the unit disc.

2.6 Moment noise sensitivity

Various invariant moment schemes have proved useful in recognition and reconstruction tests [2, 17, 28, 78]. They have proved successful and have shown invariance properties for images containing very little or no noise. However in the presence of noise, the computed Hu invariant moments M_{1-7}, begin to degrade. One study [80] showed that higher order moments are more vulnerable to white noise, thus making their use undesirable for pattern recognition. A more recent study [70] compared the performance of the Hu invariant moments with a set of moments based on wavelet basis functions. This study showed that when using Hu's moments, even a slight discrepancy in the image can cause considerable confusion (i.e. minor shape deformation or digitisation errors) if trying to discriminate between two similar images. However, noise simulation (in terms of image analysis) is very involved, and is highly dependent on the type of noise being simulated, its distribution, how it is applied etc. These noise-related issues are further discussed in Sections 6.3 and 6.7. It must be noted that while studies involving noise analyses may be correct for each specific test condition, care must be taken when generalising to alternative noise-related conditions.

2.7 Conclusions

This chapter has detailed the conventional use of statistical moments - the analysis of single two-dimensional images. Non-orthogonal and orthogonal descriptors have been discussed, describing moments which possess various useful properties including translation, scale and rotation invariance. We have considered both image description and reconstruction. These techniques are applied to single images and describe a shape in terms of its spatial (or pixel) distribution. However, many computer vision and image processing problems involve the analysis of image sequences. For example, analysing the movement of an aircraft in the sky, or ultra-sound images of a beating heart. These image sequences can be of a rigid shape (i.e. an aircraft) or a deforming one (i.e. a beating heart), while consecutive images within the sequence tend to be highly correlated. Thus, a general framework called velocity moments has been developed to utilise the useful properties of statistical moments enabling the analysis of moving features within image sequences. The next chapter introduces the structure of these moments, here we are primarily interested in their description capability, although the problem of reconstruction is also covered.

Chapter 3

Velocity moments

3.1 Introduction

One method of developing a technique to analyse image sequences is to stack the images into a three dimensional XYT (x, y plus time) block, and then apply a 3D descriptor to this data. Data in this form could be described using conventional 3D moments (i.e. Cartesian 3D moments [69]), treating time as the z axis. However, this method confounds the separation of the time and space information, as they are embedded in the data. Time is fundamentally different from the spatial analysis, thus, we intend to acknowledge this by treating it separately. Further, we are interested in the ability of separating time and space, enabling description of motion and/or space.

An alternative method to analyse image sequences is to reformulate the descriptor to incorporate time, enabling the separation of the time and spatial descriptions (if desired), resulting in a more versatile descriptor. To achieve this, a method of motion description within the moment basis is required. We have already seen how the COM (centre of mass) describes a unique global position within the field of view forming the basis for the centralised moments (Section 2.2.2). Using the COM descriptions between consecutive images, a description of global motion in either axis is possible. Further, the COM is guaranteed to exist, independent of the distribution (unlike alternative higher order moments), justifying the use of this low order moment as the basis of a generic framework.

Our new velocity moments are based around the COM description and are primarily designed to describe a moving and/or changing shape in an image sequence. The method enables the structure of a moving shape to be described, together with motion information. The velocity moments are calculated from a sequence of

images. Their generalised structure is:

$$A_{mn\mu\gamma} = \sum_{i=2}^{I} \sum_{x} \sum_{y} U \; S \; P_{i_{xy}} \tag{3.1}$$

Here the shape's structure (from each image i) contributes through each pixel $P_{i_{xy}}$ and the weighting function S. Here S is either a centralised Cartesian polynomial [28], or a Zernike polynomial [87]. Motion, or velocity, is introduced through U as the differences between consecutive COMs in the image sequence. The Cartesian monomials were first studied due to their simplicity and ease of computation. Further to this, the orthogonal Zernike moments are a well established and proven standard technique (in both image noise and pattern recognition), providing an ideal platform to enable the analysis of the new framework on an orthogonal set.

3.2 Cartesian velocity moments

The Cartesian velocity moments [74] are computed from a sequence of images as:

$$vm_{pq\mu\gamma} = \sum_{i=2}^{I} \sum_{x=1}^{M} \sum_{y=1}^{N} U(i,\mu,\gamma) \; S(i,p,q) \; P_{i_{xy}} \tag{3.2}$$

where $S(i,p,q)$ arises from the centralised moments:

$$S(i,p,q) = (x - \overline{x_i})^p (y - \overline{y_i})^q \tag{3.3}$$

and $U(i,\mu,\gamma)$ introduces velocity as:

$$U(i,\mu,\gamma) = (\overline{x_i} - \overline{x_{i-1}})^\mu (\overline{y_i} - \overline{y_{i-1}})^\gamma \tag{3.4}$$

$\overline{x_i}$ is the current COM in the x direction, while $\overline{x_{i-1}}$ is the previous COM in the x direction, $\overline{y_i}$ and $\overline{y_{i-1}}$ are the equivalent values for the y direction. (The image sequence is assumed to begin at image index $i = 1$, however, summation commences at $i = 2$ to ensure that the first velocity calculation $U(2,\mu,\gamma)$ is defined, essentially achieving invariance to the start position within the field of view). It can be seen that the equation can easily be decomposed into averaged centralised moments (vm_{1100}), and then further into an averaged Cartesian moment (vm_{1100} with $\overline{x_i} = \overline{y_i} = 0$). The velocity moments for which $\mu = \gamma = 0$ are:

$$vm_{pq00} = \sum_{i=2}^{I} \sum_{x=1}^{M} \sum_{y=1}^{N} (x - \overline{x_i})^p (y - \overline{y_i})^q P_{i_{xy}} \tag{3.5}$$

which are the averaged centralised moments. Setting $p = q = 0$ produces:

$$vm_{00\mu\gamma} = \sum_{i=2}^{I}\sum_{x=1}^{M}\sum_{y=1}^{N}(\overline{x_i} - \overline{x_{i-1}})^{\mu}(\overline{y_i} - \overline{y_{i-1}})^{\gamma}P_{i_{xy}} \tag{3.6}$$

which is a summation of the difference between COMs of successive images (i.e. the velocity). The structure of Equation 3.2 allows the image structure to be described together with velocity information from both the x and y directions. These results are averaged by normalising with respect to the number of images and the average area of the object. This results in pixel values for the velocity terms, where the velocity is measured in pixels per image. The normalisation is expressed as:

$$\overline{vm_{pq\mu\gamma}} = \frac{vm_{pq\mu\gamma}}{A(I-1)} \tag{3.7}$$

where A is the average area (number of pixels) of the moving object, I is the number of images and $\overline{vm_{pq\mu\gamma}}$ is the normalised Cartesian velocity moment.

3.2.1 Reconstruction

Due to the non-orthogonal monomials that they are based on, the Cartesian velocity moments are also non-orthogonal, which suggests that the moment matching reconstruction method is applicable. If we consider the velocity moment case for a single image, then this is actually a centralised moment for a single image, so the method described in Section 2.2.3 holds. To consider the case of the image index i, for $I > 2$, the problem can be simplified by assuming that the resulting velocity moments describe a single image, even though they are in fact derived from a sequence of images. For example, the velocity moments computed from a sequence of a rigid moving shape would effectively produce a refined description of the rigid shape, averaging out the affects of any noise in the sequence. To consider attempting to reconstruct the complete image set, the COM's for each separate image would be needed, along with the velocity moment value before it is summed and averaged over the complete sequence. So to reproduce a single combined image, $g(x, y)$ (Equation 2.23) can be defined as:

$$g(x,y) = \sum_{j,k,l,m=0}^{N_{max}} g_{jklm}(x-\overline{x})^j(y-\overline{y})^k a^l b^m \tag{3.8}$$

where:

$$a = average\ x\ velocity\ =\ vm_{0010}\ ,\ \ b = average\ y\ velocity\ =\ vm_{0001} \tag{3.9}$$

Here $\overline{x}$ and $\overline{y}$ are assumed to be the averaged COM's in the x and y directions. Similarly Equation **??** becomes:

$$\int_{-1}^{1}\int_{-1}^{1} g(x,y)(x-\overline{x})^j(y-\overline{y})^k a^l b^m \; dx \; dy \equiv vm_{jklm} \tag{3.10}$$

Applying this in practice means that an increasing number of large CLEs need to be solved, even if only low order ($(j+k) < 4$) reconstruction is required. However it would appear that an order of at least $(j+k) \gg 8$ is needed to produce a meaningful reconstruction (as found in Section 2.2.3). As such the computation involved appears excessive, though it certainly would appear that the new velocity moments do allow for partial reconstruction, should it be desired. Reconstruction may prove useful to investigate which moments give rise to which characteristics within the image sequence, or vice versa. This topic is discussed briefly in Section 7.1.4.

3.2.2 Individual-image scale invariance

The centralised moments can be normalised with respect to scale using Equation 2.20. This normalisation can be applied to the Cartesian velocity moments to produce individual-image scale invariance. First we consider the simple case of two consecutive images ($I = 2$), from a sequence of a moving and rotating object with constant shape. Applying this to the Cartesian velocity moments produces:

$$\begin{aligned} vm_{pq\mu\gamma} &= \sum_{i=2}^{I}\sum_{x=1}^{M}\sum_{y=1}^{N} U(i,\mu,\gamma)\ (x-\overline{x_i})^p(y-\overline{y_i})^q\ P_{i_{xy}} \\ &= U(2,\mu,\gamma)\ \sum_{x=1}^{M}\sum_{y=1}^{N}(x-\overline{x_2})^p(y-\overline{y_2})^q\ P_{2_{xy}} \end{aligned} \tag{3.11}$$

$U(2,\mu,\gamma)$ is scalar and the remainder of the expression is just a centralised moment, the result can be normalised with respect to scale using Equation 2.20 producing:

$$\begin{aligned} vm_{pq\mu\gamma} &= U(2,\mu,\gamma)\frac{\sum_{x=1}^{M}\sum_{y=1}^{N}(x-\overline{x_2})^p(y-\overline{y_2})^q\ P_{2_{xy}}}{\mu_{00}^{\gamma}} \\ &= U(2,\mu,\gamma)\ \eta_{pq} \end{aligned} \tag{3.12}$$

If a sequence of images with $I > 2$ is considered, then the Cartesian velocity moments need to be reformulated to include (individual image) scale invariance:

$$vm_{mn\mu\gamma} = \sum_{i=2}^{I} U(i,\mu,\gamma)\left(\frac{\sum_{x=1}^{M}\sum_{y=1}^{N}(x-\overline{x})^m(y-\overline{y})^n\ P_{i_{xy}}}{\mu_{i_{00}}^{\gamma}}\right)$$

$$= \sum_{i=2}^{I} U(i, \mu, \gamma)\ \eta_{i_{mn}} \tag{3.13}$$

subject to Equation 2.21, where $\mu_{i_{00}}$ is the i^{th} zero-order moment (μ_{00}) and $\eta_{i_{mn}}$ is the i^{th} scale normalised centralised moment, Equation 2.20. It must be noted that this scale invariance will increase the between-image correlation, which in turn increases the already highly correlated (non-orthogonal) description. However, this modification to the descriptor may be useful for specific applications, such as describing an object which is moving towards the camera. One direct application could be, say, providing a feature to follow at a football match to allow the cameras to focus on the action. Equally individual-image scale invariance is also possible by rescaling, or normalising the image prior to calculation, a technique explained in Section 2.4.2 and used later in Section 3.3.

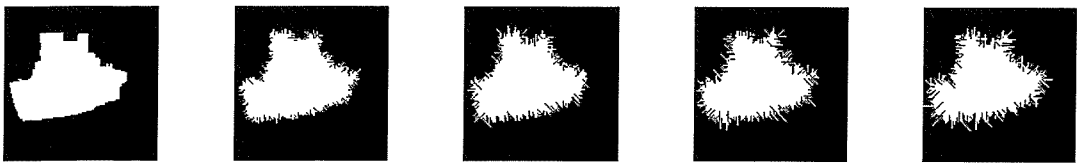
3.2.3 Simple moving shape recognition and perimeter noise

In order to assess the performance of the Cartesian velocity moments on extracted images, tests were run on synthetic images. By applying the velocity moments to three different sequences the recognition capabilities were examined. The sequences were of a square and triangle moving along the x axis at 5 pixels/image and a circle moving at 7 pixels/image. The circle also had a small movement of 0.1 pixels/image in the y axis. Each sequence consisted of ten images. These produced significantly different second order moments, Table 3.1 (see Section 3.2.4 for an explanation of what these moments represent), and the x velocity term vm_{0010} was estimated correctly in each case.

Index	Square	Triangle	Circle
$\overline{vm_{2201}}$	0.00e00	0.00e00	1.32e04
$\overline{vm_{2010}}$	0.16e04	6.65e02	0.26e04
$\overline{vm_{2210}}$	0.55e06	2.66e04	0.74e06
$\overline{vm_{0010}}$	5.00e00	5.00e00	7.00e00

Table 3.1: Low order simple moving shape Cartesian velocity moments

A common problem in the application of a new technique to a real-world situation is the issue of image noise - a topic which will be frequently re-visited throughout this thesis. Applying increasing amounts of noise across each image in a sequence will cause the COM calculations (for each image) to slowly drift towards the centre of the image (irrespective of the shape). This process essentially 'smears' the shape across the image. Further, the addition of such noise in an image sequence can be perceived as a pre-processing problem, as arguably 'salt and pepper' noise could be removed by a simple median filter. Due to these considerations it was decided to investigate the effects of perimeter noise on the velocity moments. This



(a) Original (c) $\sigma^2 = 2.5$ (d) $\sigma^2 = 5.0$ (e) $\sigma^2 = 7.5$ (e) $\sigma^2 = 10.0$

Figure 3.1: Original tug-boat image and example perimeter noise images.

(a) Original (c) $\sigma^2 = 2.5$ (d) $\sigma^2 = 5.0$ (e) $\sigma^2 = 7.5$ (e) $\sigma^2 = 10.0$

Figure 3.2: Original overlaid-shapes image and example perimeter noise images.

was achieved using a set of synthetic image sequences with the perimeter of the shape degraded by noise, simulating poor contour extraction. To provide a basis for comparison, a set of averaged Hu [28] invariant moments was used. The central limit theorem [61] states that given a population distribution, a distribution of samples about its mean approaches a normal, or Gaussian distribution, given enough samples. The larger the number of samples, the better this approximation becomes. In this limit we can assume all noise (perimeter or otherwise) to be Gaussian distributed. Due to the higher order Hu invariant moments being close to zero for symmetrical shapes (i.e. a circle), two sequences of moving asymmetrical shapes were used, each consisting of nine images. The first consisted of a series of overlaid geometrical shapes, the second was a silhouette of a tug-boat (all images were 128×128). Figures 3.1 and 3.2 show an image from each sequence with added perimeter noise. The zero mean Gaussian distributed noise function was able to both add and remove pixels from the perimeter of the shape, effectively moving the perimeter pixels into the shape, or outwards away from their original position. (A detailed explanation of the perimeter noise algorithm can be found in Appendix A.1.1.) The amount of perimeter noise applied was adjusted using the variance of the Gaussian process. The variance took values from 0.0 pixels (no noise) to 10.0 pixels, in 0.1 steps. The performance of the two methods was then plotted and compared, with example results shown in Figure 3.3 and 3.4. The moment values have been plotted in terms of their percentage deviation from the original no-noise value. Further examples and analysis can be found in Appendix A. The motion estimates (vm_{0010}) shown in Figure 3.3 have a peak-to-peak variation of approximately 7% of the original (no noise) value, a relatively low variation given the original value was small at 3.64 pixels/image. Clearly the new velocity moments (Figure 3.4b shows vm_{2010}) are much less affected by noise, changing at most by

9 %, whereas the traditional moments can change by a great amount even when averaged, Figure 3.4a shows $\simeq$ 40 % variation for I_3. The effects of the perimeter noise on the new technique can be seen in Figure 3.4b, where the gradually increasing (mean) Cartesian velocity moment value reflects the spread of the image in the x axis, with respect to the x direction velocity. A theoretical analysis of the effects of the perimeter noise (resulting in similar conclusions) can be found in Appendix A. Here we have modelled poor extraction using a zero mean Gaussian fluctuation about the perimeter of a shape. Intuitively it will have little or no effect on the COM calculations, a conclusion which has been illustrated by these results and by theoretical analysis (also shown in Appendix A). As a direct result little effect on the motion information in the velocity moments will occur (see Figure 3.3), as these are the differences between consecutive COM calculations. However, the effects will be visible in the higher order moments, since the moments describe the image distribution, for example see Figure 3.4b. Some effects will be apparent in the low order moments due to the discrete nature of both the implementation and the approximation of a Gaussian process. By exploiting temporal correlation within the sequence, these effects can be reduced. Applying the same noise model to the averaged invariant Hu moments will have the same effect, refer to Section A.1.3. Again the correlation of the sequence can be exploited, although the effects of the noise are amplified by the non-linear combinations of the centralised moments comprising the Hu invariant set, producing results like those shown in Figure 3.4a. (These non-linear combinations are used to produce rotation invariance.) Modelling poor extraction in this manner has illustrated how the velocity moment structure incorporating motion description is less affected by uncorrelated perturbations in the shape's perimeter. Using a sequence of images can further improve the spatial descriptions, provided that non-linear combinations of the velocity moments are not employed. Finally it must be noted that the results produced by analyses like this are dependent on the image content, making them application dependent. Different moments describe different aspects of an image distribution, therefore if the characteristic produced by the noise was not present in the image previously, then the noise will understandably have a large effect. This can be illustrated by considering the moving circle image sequence. Cartesian moments of order $(p + q) > 2$ should be zero (or approximately zero) for a binary image of a circle. The addition of the perimeter noise will cause these moments to oscillate either side of zero, as they will only be affected by the noise and not by the original shape. In the same way the effects of the perimeter noise on the motion information in the velocity moments (i.e. vm_{0010}) will be proportional to the size of the motion present in the image sequence. Thus, the smaller the average motion, the greater the effect the noise will have. These effects can be described in terms of the signal-to-noise ratio, between

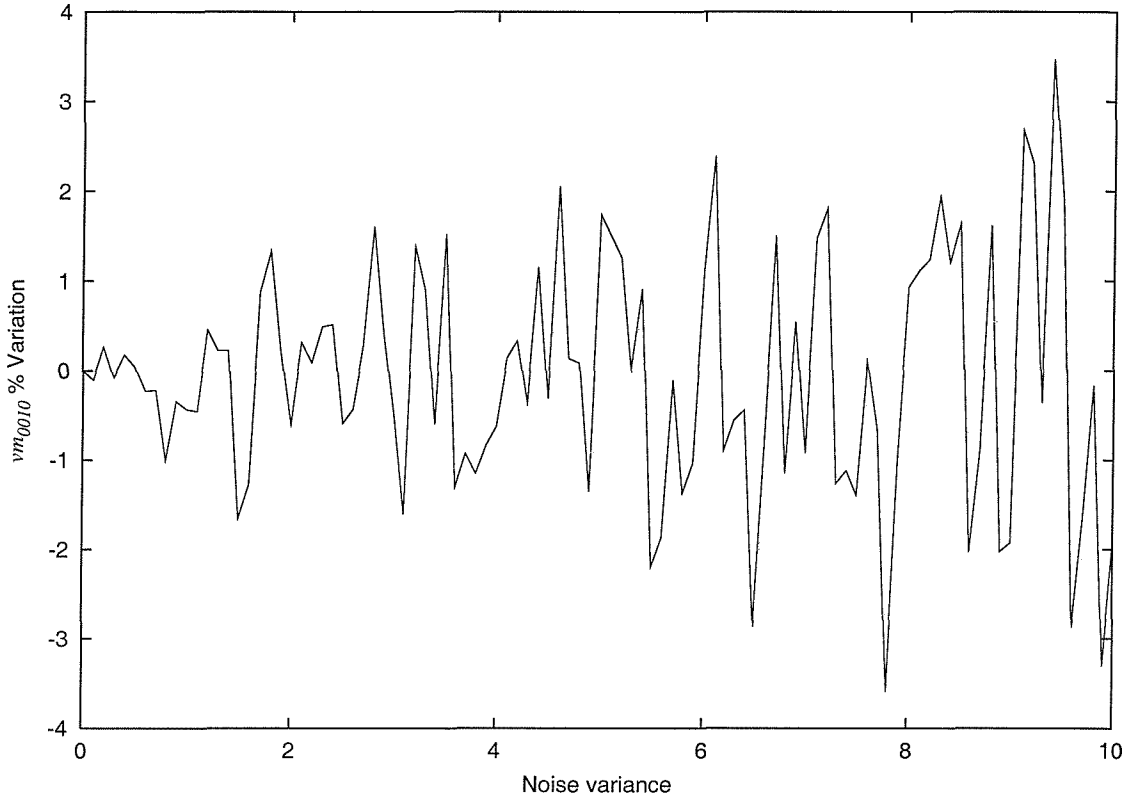


Figure 3.3: Example vm_{0010} result for the overlaid-shapes sequence against increasing perimeter noise variance.

the image content and the perimeter noise function.

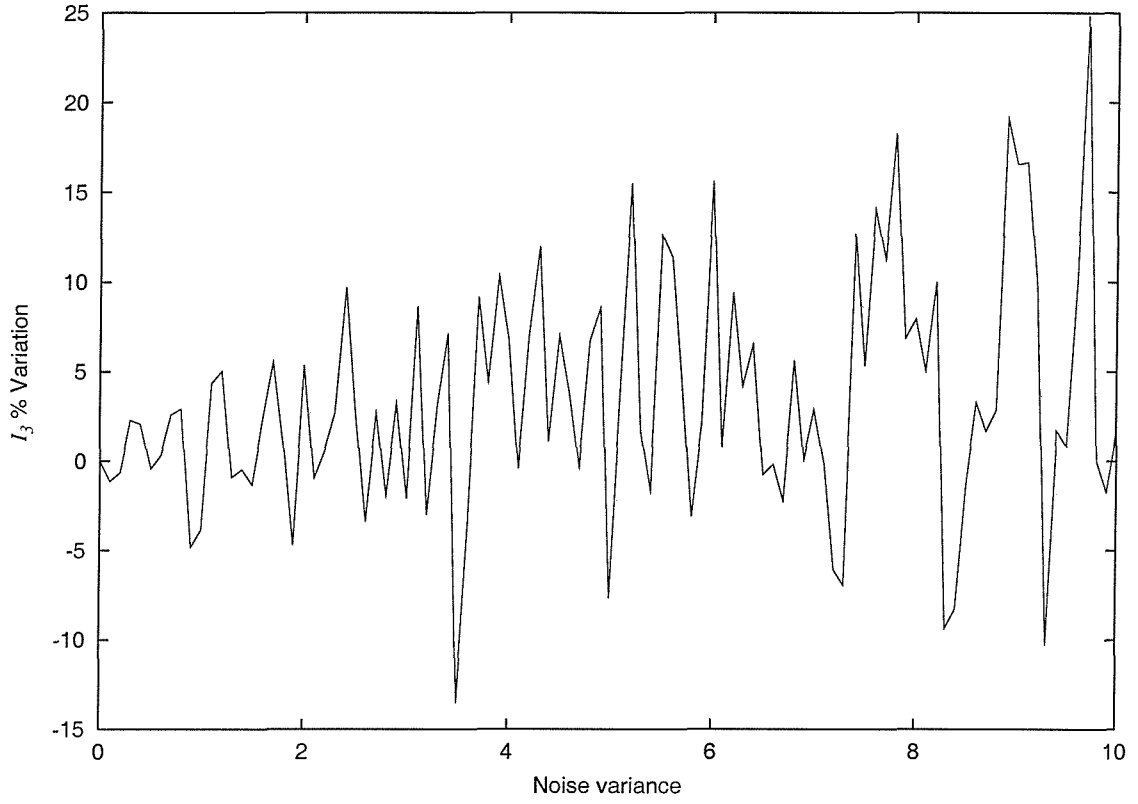
3.2.4 Understanding the low order moments

Object mean

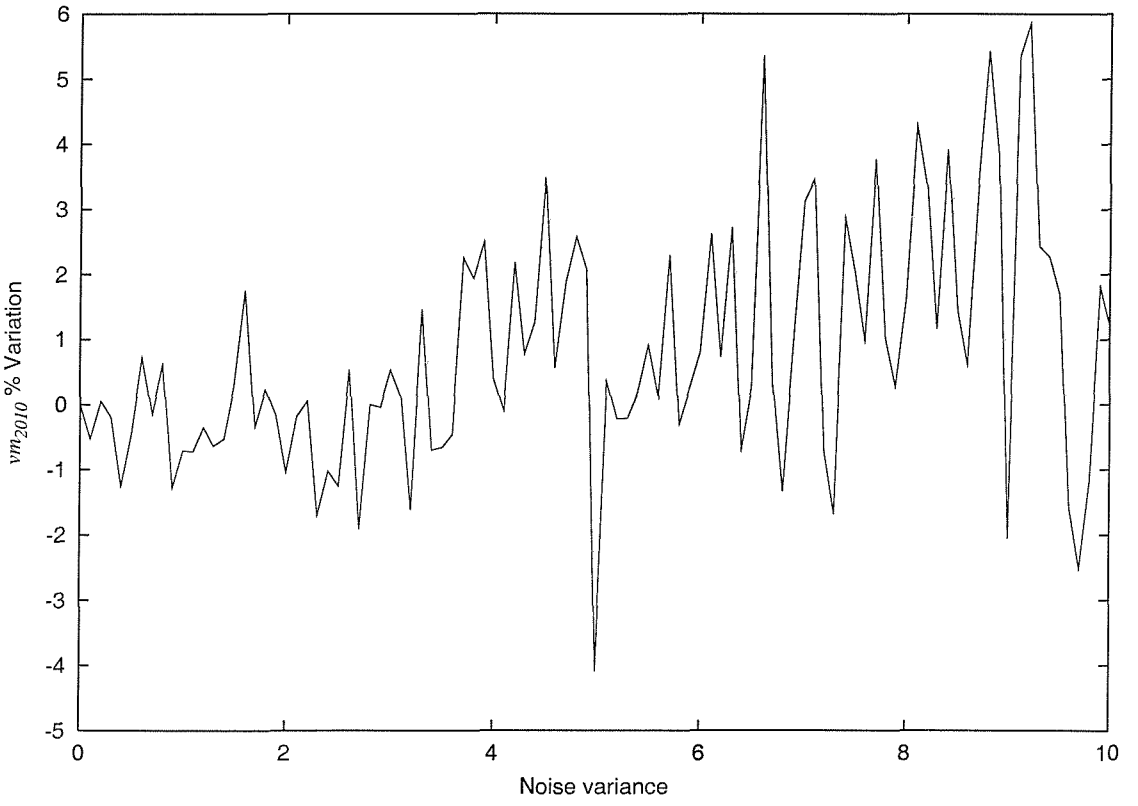
If we consider a Cartesian moment description m_{10}, then this can be considered as the mean value of the pixels in the x direction. Alternatively m_{01} is the mean value of y. As we have already shown in Section 2.2.1, these values for a binary image can be interpreted as the central co-ordinates of image. In this way it is easy to locate the centre of an single object present in the image. Following on from this, vm_{1000} is the averaged centralised moment, μ_{10}. This is effectively the averaged and translated Cartesian moment of order m_{10}. Therefore, vm_{1000} describes the pixel construction with respect to the x dimensions of the object as it moves through the sequence.

Object spread

If we now consider the second order moment m_{20}, this describes the spread of pixels in the x direction only. If calculated about the mean (i.e. the centralised moment μ_{20}), then this is the image's x direction (sample) variance. Similarly μ_{02} is the variance about the y axis and μ_{11} is the image's co-variance. Therefore, the velocity moment vm_{1100} is an expression of the image sequence's time averaged co-variance.



(a) Hu invariant moment I_3 with increasing perimeter noise variance.



(b) vm_{2010} with increasing perimeter noise variance.

Figure 3.4: Example Hu and velocity moment results for the overlaid-shapes sequence against increasing perimeter noise variance.

Using this information the velocity moment vm_{1110} contains information about the average pixel spread in both the x and y axis (about the mean) with respect to the shape's x direction velocity. It can be seen that these moments are more resilient to noise that degrades the density of the shape, due to them containing information about the spread of the object. Strictly speaking, when discussing variance calculations on an image using Cartesian moments, we are actually referring to the un-normalised sample variance, (in contrast to the population variance), as the mean value used to calculate the variance is determined from the samples themselves (i.e. the image) rather than being known *a priori* and the result has not been normalised with respect to the number of samples.

Movement orientation

The direction and orientation of the motion can then described using the moments vm_{2201} and vm_{2210}. This is possible as vm_{2200} will always be positive due to the values of p and q. By looking at the signs of these moments an estimate of the direction of motion can be determined, assuming the orientation of the image space is known *a priori*. vm_{2202} describes the average magnitude of velocity (squared) in the y direction. Similarly vm_{2220} is the average magnitude of velocity (squared) in the x direction. vm_{2200} will be the same irrespective of the direction of movement, since the centralised moments are invariant to translation. Essentially the shape can move along any orientation, without any effect on the moment value. In practice there will be slight variation in the moment value due to the discrete implementation. Neglecting this, the magnitude of the velocity term is therefore invariant with direction and orientation.

3.3 Zernike velocity moments

The new Zernike velocity moments are expressed as:

$$A_{mn\mu\gamma} = \frac{m+1}{\pi} \sum_{i=2}^{I} \sum_{x} \sum_{y} U(i,\mu,\gamma)\ S(m,n)\ P_{i_{xy}} \tag{3.14}$$

They are bounded so that $(x^2 + y^2) \leq 1$, while the shape's structure contributes through the orthogonal polynomials:

$$S(m,n) \ = \ [V_{mn}(r,\theta)]^* \tag{3.15}$$

Velocity is introduced as before (Equation 3.4), while normalisation is produced by:

$$\overline{A_{mn\mu\gamma}} = \frac{A_{mn\mu\gamma}}{A(I-1)} \tag{3.16}$$

The coordinate values for $U(i, \mu, \gamma)$ are calculated using the Cartesian moments and then translated to polar coordinates. If we consider first the x direction case only, from Equation 2.56 the angle θ for a difference in x position is either 0 or π radians. The value used is dependent on the direction of movement. If the movement is left to right then:

$$x = r\ \cos\theta = r\ \cos(0) = r \tag{3.17}$$

where r is the length of the vector from the previous COM to the current COM, i.e. the velocity in pixels/image. Alternatively, if the movement is right to left then:

$$x = r\ \cos\theta = r\ \cos(\pi) = -r \tag{3.18}$$

The mapping to polar coordinates results in a sign change that could be used to detect the direction of motion. Similarly for the y direction velocity, the values of θ are either $\frac{\pi}{2}$ or $\frac{3\pi}{2}$ radians, and using Equation 2.56 produces:

$$y = r\ \sin\theta = r\ \sin\left(\frac{\pi}{2}\right) = r \tag{3.19}$$

and

$$y = r\ \sin\theta = r\ \sin\left(\frac{3\pi}{2}\right) = -r \tag{3.20}$$

3.3.1 Orthogonality condition

The normalised orthogonality condition for the Zernike moments is [53]:

$$\int_0^{2\pi}\int_0^1 V_{nl}^*(r,\theta)V_{mk}(r,\theta)\ r\ dr\ d\theta = \frac{\pi}{n+1}\delta_{nm}\delta_{lk} \tag{3.21}$$

where:

$$\delta_{ab} = 1 \text{ for } a = b,\ \delta_{ab} = 0 \text{ for } a \neq b \tag{3.22}$$

where here the Zernike polynomial (or basis function) $V_{mk}(r,\theta)$, is being analysed. Again we consider the simple case of two consecutive images ($I = 2$) from a sequence of a moving shape. Applying the Zernike velocity moments to this sequence enables one image to be described (the second of the two images), along with the motion information from between the two images. From Equation 3.14 the basis function of the Zernike velocity moments is $U_{ab}(i,\mu,\gamma)V_{ab}(r,\theta)$, using this and Equation 3.21 produces the unnormalised orthogonality condition (for $I = 2$ and):

$$\begin{aligned} & \int_0^{2\pi}\int_0^1 [U_{nl}(2,\mu,\gamma)V_{nl}^*(r,\theta)]\,[U_{mk}(2,\mu,\gamma)V_{mk}(r,\theta)]\ r\ dr\ d\theta \\ = \quad & U_{nl}(2,\mu,\gamma)U_{mk}(2,\mu,\gamma)\int_0^{2\pi}\int_0^1 V_{nl}^*(r,\theta)V_{mk}(r,\theta)\ r\ dr\ d\theta \end{aligned}$$

$$= \quad U_{nl}(2,\mu,\gamma)U_{mk}(2,\mu,\gamma)\frac{\pi}{n+1}\delta_{nm}\delta_{lk} \tag{3.23}$$

where Equation 3.22 holds and $U_{ab}(i,\mu,\gamma)$ is a real-valued scalar subject to:

$$U_{ab}(i,\mu,\gamma) > 0 \tag{3.24}$$

If the shape is constant, such as a car and more than two images are processed, then each individual image's spatial descriptions remain orthogonal, just weighted by $U_{ab}(i,\mu,\gamma)$, Equation 3.23. However, the overall description of the sequence becomes correlated, due to the high similarity between the images. Further, if we consider Zernike velocity moments describing just spatial information (no motion), then the resultant temporal correlation is exploited, refining the description of the rigid shape as the sequence progresses. The final Zernike velocity moments can be considered as refined (or averaged) Zernike moments of a single image, the descriptions of which are orthogonal (Equation 3.21). If the shape is both moving and deforming (non-rigid), such as a person walking, then the correlation between consecutive image descriptions is reduced. However, this correlation is the result of the temporal sequence and is advantageous. Increasing the size of the image sequence refines the description of the moving shape within it.

In conclusion the Zernike velocity moments are a weighted sum of the Zernike moments over multiple consecutive images. The weighting is real-valued and scalar, therefore the spatial description of each consecutive image in the sequence remains orthogonal. However, the overall description is temporally correlated due to the images being a consecutive temporal sequence.

3.3.2 Reconstruction

Here we assume that the Zernike velocity moments produced describe a single image, even though they are in fact derived from a sequence of images, (as previously described for the Cartesian velocity moment case in Section 3.2.1). To attempt to recreate each image within the sequence would require the COM values for each separate image, along with the Zernike moment value before they were summed. Whereas if a single image is to be reconstructed from the sequence of a rigid shape, then the moments produced are effectively a refined description of the moving shape over time. So if applied to the moments describing a rigid shape moving in the presence of noise, the velocity moments will average out the effects of that noise. Assuming we are describing a rigid shape, then the orthogonality condition holds (Section 3.3.1), as each single image's spatial descriptions remain orthogonal. The overall description of the rigid shape is refined due to the highly correlated sequence description. Equation 2.66 can be extended using the Zernike velocity moments, shown here for the $\mu = \gamma = 1$ case:

$$\hat{f}(r,\theta) = \sum_{m=0}^{N_{max}} \sum_{n} A_{mn11} V_{mn}(r,\theta) \tag{3.25}$$

where n is constrained by Equation 2.50, expanding this expression (as before in Section 2.4.3) using real-valued functions, produces:

$$\hat{f}(r,\theta) = \sum_{m=0}^{N_{max}} \sum_{n>0} (C_{mn11} \cos n\theta + S_{mn11} \sin n\theta)\ R_{mn}(r) + \frac{C_{m011}}{2} R_{m0}(r) \tag{3.26}$$

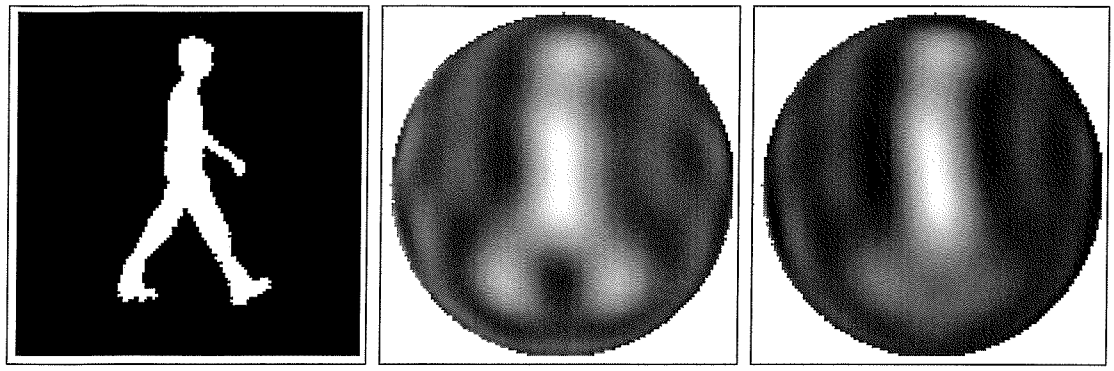
where:

$$\begin{aligned} C_{mn11} &= \frac{2Re\left[A_{mn11}\right]}{(I-1)} \\ &= \frac{2m+2}{(I-1)\ \pi} \sum_{i=2}^{I} \sum_{x} \sum_{y} U(i,1,1)\ f_i(r,\theta) R_{mn}(r) \cos n\theta \qquad (3.27) \\ S_{mn11} &= \frac{-2Im\left[A_{mn11}\right]}{(I-1)} \\ &= \frac{-2m-2}{(I-1)\ \pi} \sum_{i=2}^{I} \sum_{x} \sum_{y} U(i,1,1)\ f_i(r,\theta) R_{mn}(r) \sin n\theta \qquad (3.28) \end{aligned}$$

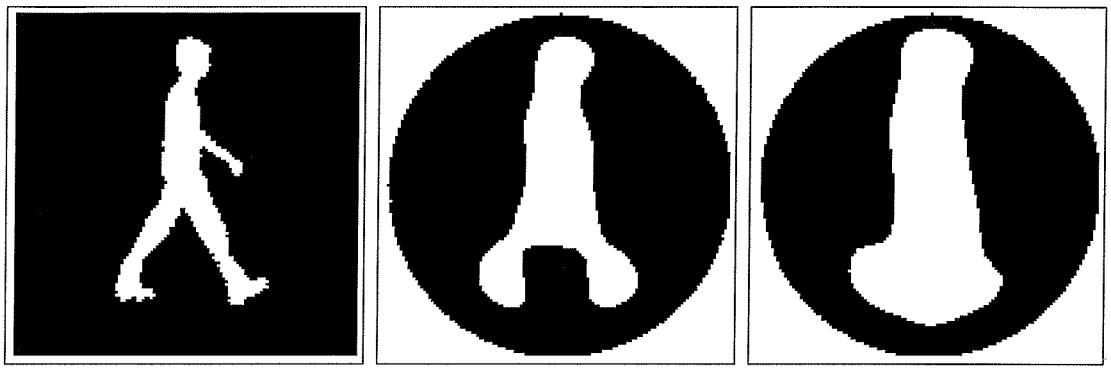
and each image is bounded by $(x^2+y^2) \leq 1$. Figure 3.5c shows an example Zernike velocity moment reconstructed image (orders $2-12$ were used for reconstruction). The Zernike velocity moments were generated from the consecutive (binary silhouette) sequence of a person walking. An example image from the sequence (showing the person mid-stride) is displayed in Figures 3.5a and d. The reconstructed results reflect the time-averaged nature of the velocity moments, producing an image which blurs the area containing leg motion. This blurred contour becomes clearer when thresholded, as shown in Figure 3.5f. While the person's torso, (whose position and motion will be more constant throughout the image sequence) appears well represented. Evidence of Gibbs phenomena [75] is also visible in the un-thresholded image. In comparison the (single image) reconstructed version of the same person mid-stride can be seen in Figure 3.5b, (constructed using Figure 3.5a as the source). While the reconstructed version is clearly lacking in detail (or high frequency information), the separation between the person's legs is apparent, unlike its sequence constructed version (Figures 3.5c and f).

3.3.3 Rotation invariance

We have already seen that the absolute value of a Zernike moment is rotation invariant as reflected in the mapping of the image to the unit disc, Section 2.4.2. Re-iterating this, the rotation of the shape around the unit disc is expressed as a phase change, if ϕ is the angle of rotation, A^R_{mn} is the Zernike moment of the rotated



(a) Example silhouette. (b) Single image reconstruction. (c) Sequence reconstruction.



(d) Example silhouette. (e) Thresholded (b). (f) Thresholded (c).

Figure 3.5: Zernike velocity moment reconstruction (order $2-12$) example.

image and A_{mn} is the Zernike moment of the original image then:

$$A_{mn}^{R} = A_{mn} \exp(-jn\phi) \tag{3.29}$$

To apply this to the Zernike velocity moments, we again consider the simple case of two consecutive images ($I = 2$). However, here the rigid shape is moving and rotating object. Calculating the magnitude of the Zernike velocity moment produces rotation invariance, (assuming the shape is moving at the same spatial velocity, although the speed of rotation can vary). Here the description includes velocity information (between the images), while describing the structure of the shape from the second image, essentially a velocity-weighted Zernike moment:

$$\begin{aligned} A_{mn\mu\gamma} &= \frac{m+1}{\pi} \sum_{i=2}^{I} \sum_{x} \sum_{y} U(i,\mu,\gamma)\ S(i,m,n)\ P_{i_{xy}} \\ &= U(2,\mu,\gamma)\ \frac{m+1}{\pi} \sum_{x} \sum_{y} S(2,m,n)\ P_{2_{xy}} \\ &= U(2,\mu,\gamma)\ A_{mn} \end{aligned} \tag{3.30}$$

since rotation introduces a phase change (Equation 2.60) then:

$$\begin{aligned} &= U(2,\mu,\gamma)\ A_{mn}\exp(-jn\phi) \\ &= U(2,\mu,\gamma)\ A^R_{mn} \end{aligned} \tag{3.31}$$

where again, ϕ is the angle of rotation, A_{mn} is the original Zernike moment (unrotated and stationary) and $U(2,\mu,\gamma) > 0$. The magnitude $\mid U(2,\mu,\gamma)\ A^R_{mn} \mid$ will be rotation invariant. A longer sequence of images ($I > 2$) will not maintain rotation invariance, since:

$$\mid A_{mn} \mid + \mid A_{mn} \mid \ \neq \ \mid 2A_{mn} \mid \tag{3.32}$$

However, through a slight modification, rotation invariance for $I > 2$ is possible. The rotation invariant Zernike velocity moment $R_{mn\mu\gamma}$ is:

$$\begin{aligned} R_{mn\mu\gamma} &= \sum_{i=2}^{I} U(i,\mu,\gamma)\ \mid \left[\frac{m+1}{\pi}\sum_x\sum_y S(i,m,n)P_{i_{xy}}\right] \mid \\ &= \sum_{i=2}^{I} U(i,\mu,\gamma) \mid A_{i_{mn}} \mid \end{aligned} \tag{3.33}$$

as bounded by $(x^2+y^2) \leq 1$ and where $A_{i_{mn}}$ is the i^{th} Zernike moment. A simple experiment can be used to demonstrate the rotation properties of Equation 3.33. Tables 3.2 and 3.3 show the results of applying this description to multiple sequences of moving and rotating images (128×128) of the character 'A' (the moment values are normalised with respect to sequence length and mass using Equation 3.16). Table 3.2 shows moments which describe purely spatial information, while Table 3.3 shows moments which include both spatial and motion information. Four sequences were tested, where each sequence had a different angle of rotation (between consecutive images), from 0^o through to 90^o rotation, producing different speeds of rotation. Figure 3.6 shows five example consecutive images from the 30^o sequence. Each sequence consisted of thirteen images of the rotating character moving at 2 pixels per image in the x direction. The direction of rotation was anti-clockwise for all sequences except the 90^o case, where the rotation was clockwise. Tables 3.2 and 3.3 show the corresponding sample mean μ, standard deviation σ, and $\sigma/\mu\,\%$ (coefficient of variation), indicating the percentage spread of the moment values. Small values of $\sigma/\mu\,\%$ indicate a compact set (or cluster) of moments. Table 3.3 shows example motion versions of those velocity moments shown in Table 3.2. (For example, the values for R_{2010} are approximately double that of the values for R_{2000}. This is due to R_{2010} being effectively a time averaged $2R_{2000}$ due to the constant x direction motion of 2 pixels per image.) It can be seen that $\sigma/\mu\,\%$ is 0.94 for both R_{2000} and R_{2010}, while all of the results have values of $\sigma/\mu\,\% < 7$.

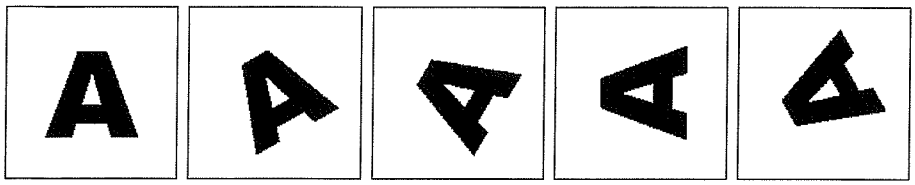


Figure 3.6: Consecutive windowed images from the 30^o rotation sequence.

Rotation sequence	$\overline{R_{2000}}$	$\overline{R_{2200}}$	$\overline{R_{3100}}$	$\overline{R_{3300}}$
0^o	2190.07	98.21	162.79	527.93
30^o	2170.34	105.46	166.36	552.41
60^o	2169.20	105.34	166.36	553.27
90^o	2140.73	110.02	188.05	563.10
μ	2167.59	104.76	170.89	549.18
σ	20.31	4.88	11.56	14.97
$\sigma/\mu\%$	0.94	4.66	6.77	2.73

Table 3.2: Purely spatial rotation invariant descriptions.

Rotation sequence	$\overline{R_{2010}}$	$\overline{R_{2210}}$	$\overline{R_{3120}}$	$\overline{R_{3320}}$
0^o	4380.14	196.41	651.19	2111.74
30^o	4340.69	210.93	665.44	2209.62
60^o	4338.40	210.69	665.45	2213.08
90^o	4281.47	220.05	752.18	2252.42
μ	4335.18	209.52	683.57	2196.72
σ	40.61	9.77	46.23	59.88
$\sigma/\mu\%$	0.94	4.66	6.76	2.73

Table 3.3: Spatial and velocity rotation invariant descriptions.

These results are comparable with those previously calculated on single rotated images, [38, 37]. By analysing a temporal image set of a shape via Equation 3.33, the description of the moving shape can be refined, while also containing velocity and direction of movement information. Further improvements in performance would be apparent (over describing a single image) where occlusion or image noise is present. The descriptions produced appear to be both invariant to rotation and direction/speed of rotation while the addition of the velocity does not appear to affect the spread (coefficient of variation, $\sigma/\mu\%$) of the moment values. It must be noted that exact invariance is not obtained as a result of the errors introduced by the discrete implementation, both in the calculation of the moments and the mapping of the image to the unit disc. Further, the results are very much dependent on the shape itself, along with the image size. Greater errors will appear where high frequency information is present, i.e. corners within the images. Descriptions using Equation 3.33 will be more correlated than the non-rotation invariant version,

Equation 3.14. This is directly linked to the spatial description being the absolute value of the Zernike moment for each image (in Equation 3.33), rather than being composed of the respective real and imaginary parts.

For traditional Zernike moments and $R_{mn\mu\gamma}$, the effects of the rotation are very much dependent on the size of the object being rotated. This means that the scale and translation mapping can drastically alter the rotation performance, Section 2.4.2. Pixels which are closer to the edge of the unit disc will be described more efficiently, than those pixels which are grouped around the origin (centre) of the unit disc. Equally, objects with pixels falling close to the edge of the unit disc will have descriptions which are more correlated, due to the converging radial polynomials as r approaches unity, Figure 2.6.

3.4 Relating Zernike and Cartesian velocity moments

Section 2.5 explained how traditional Zernike and Cartesian moments are related, allowing conversion between the two, (i.e. expressing Zernike moments as a summed combination of Cartesian moments, or vice versa). Here this theory is applied to the two versions of the velocity moments, considering an image sequence of a moving shape. To allow comparison between the two traditional moments, the image has to be confined to $[-1, 1]$. The moving shape can be mapped to this domain using the Zernike mapping, Equation 2.58, while the motion information ($U(i, \mu, \gamma)$) is determined before this operation. The mapping sets $\overline{x_i} = \overline{y_i} = 0$ while care must be taken to ensure all of the shape is encompassed by $[-1, 1]$. The shape will appear central to the coordinate system, with respect to its mass. This in turn means that the centralised moments of the image decompose to the Cartesian moments, due to $\overline{x_i} = \overline{y_i} = 0$. If the image sequence is now described via the Cartesian velocity moments:

$$vm_{pq\mu\gamma} = \sum_{i=2}^{I} \sum_{x=1}^{M} \sum_{y=1}^{N} U(i, \mu, \gamma)\ (x - \overline{x_i})^p (y - \overline{y_i})^q P_{i_{xy}} \tag{3.34}$$

confining x, y to $[-1, 1]$ and applying the mapping produces:

$$vm_{pq\mu\gamma} = \sum_{i=2}^{I} \sum_{x} \sum_{y} U(i, \mu, \gamma)\ x^p y^q P_{i_{xy}} \tag{3.35}$$

which can be rewritten as:

$$vm_{pq\mu\gamma} = \sum_{i=2}^{I} U(i, \mu, \gamma) \sum_{x} \sum_{y} x^p y^q P_{i_{xy}} \tag{3.36}$$

Similarly the Zernike velocity moments:

$$A_{mn\mu\gamma} = \frac{m+1}{\pi} \sum_{i=2}^{I} \sum_{x} \sum_{y} U(i,\mu,\gamma)\ [V_{mn}(x,y)]^*\ P_{i_{xy}} \tag{3.37}$$

bounded by:

$$(x^2 + y^2) \leq 1 \tag{3.38}$$

can be rewritten as:

$$A_{mn\mu\gamma} = \sum_{i=2}^{I} U(i,\mu,\gamma) \frac{m+1}{\pi} \sum_{x} \sum_{y} [V_{mn}(x,y)]^*\ P_{i_{xy}} \tag{3.39}$$

The right hand side of Equations 3.36 and 3.39 are summed versions (over the images of the sequence) of the Cartesian and Zernike moments, weighted by the velocity of the moving shape. Perceived in this way, conversion between the two descriptions appears possible using the theory previously explained in Section 2.5. This analysis highlights a useful connection between the two velocity moment implementations. The Zernike velocity moments have improved characteristics over the Cartesian implementation (which is also true for traditional Zernike and Cartesian moments). However, by only using the combinations of Cartesian velocity moments which comprise their Zernike version, simple selection of less correlated features appears possible. Viewed in an alternative manner, this introduces a method of mapping the Cartesian velocity moments to an orthogonal basis, with respect to each image description. Further, this shows that it is possible to implement the Zernike velocity moments via the Cartesian coordinate system thus reducing the complexity of the calculation. It must be noted that this assumes that the Cartesian coordinate system's origin is located at the bottom left of the image. Otherwise, care must be taken when considering the direction of motion, as the signs will be different between the Cartesian and Zernike implementations. This is due to the directional information for the Zernike moments being independent of the image coordinate system, refer to Equations 3.17 through 3.20.

3.5 Scale, frame rate and sequence length invariance

Features which are independent of frame rate may be required. For example, this allows features generated on the European PAL (Phase Alternation Lines) image systems (at 25 frames/second) to be comparable to features generated on the American, Mexican, Canadian and Japanese NTSC (National Television Standards Code) systems (at 30 frames/second). A further example is the use of time-lapse filming in security applications, a technique which is investigated in Section 6.6. Normalising

the velocity moments with respect to the number of images, camera frame rate and scale is achieved by:

$$\hat{A}_{mn\mu\gamma} = \frac{A_{mn\mu\gamma}\ t_s}{A\ (I-1)} \tag{3.40}$$

while

$$t_s = \frac{1}{f_r(s)} \tag{3.41}$$

where $t(s)$ is the time between consecutive images, in seconds and $f_r(s)$ is the frame rate in seconds, (i.e. 25 frames/second). A is the average area of the moving shape and I is the number of images.

3.6 Comparison of techniques

This brief section compares the two velocity moment techniques, Cartesian and Zernike, in terms of their descriptor properties. Table 3.4 summarises these properties allowing comparison between the two implementations. Here a $\surd$ signifies that the descriptor has the corresponding property and a $*$ signifies that descriptions with the corresponding property are possible through a slight modification to the velocity moment equations, (as detailed in previous sections in this chapter). These modifications invariably produce descriptions which are more restricted, or correlated than their original versions. It can be seen that both techniques have positive attributes. Overall, the Zernike velocity moments appear more versatile. However, it must be noted that the Cartesian implementation is computationally simple and inexpensive, although simple conversion between the two techniques is possible.

Property	**Cartesian**	**Zernike**
Coordinate system	Cartesian	polar
Orthogonal (individual images)	$\times$	$\surd$
Reconstruction	$\surd$	$\surd$
Invariance property	**Cartesian**	**Zernike**
Rotation	$\times$	$*$
Translation	$\surd$	$\surd$
Scale (individual image)	$*$	$\surd$
Scale (sequence)	$\surd$	$\surd$

Table 3.4: Comparison of properties of the two velocity moment implementations.

3.7 Exploiting velocity

By including velocity in the shape description, similar objects moving in different manners can be separated. For example, we can consider two cases, a bouncing ball and a ball rolling across the floor. Traditional shape descriptors that analyse the ball itself (and are scale invariant) encode information about its shape and

structure. However distinctions between types of motion would not be addressed. For a football match, say, the commentator may be interested in what percentage of the match time the ball has spent being kicked, or moving towards one team's end of the pitch. To illustrate this point, a set of experiments were run using the Zernike velocity moments on a small database of images of a moving basketball. Three different sequences were analysed, all of which were captured on the same day. The first sequence was of the ball falling vertically (sequence 1). The second has the ball bouncing up and across the field of view (sequence 2), while in the last sequence the ball was being thrown across the field of view in a slight arc (sequence 3). The basketball in each of the sequences was then extracted using a statistical-based background subtraction technique, detailed in Appendix B. Figure 3.7 shows example images from the three original sequences and the corresponding extracted versions. The extracted images were then binary thresholded to remove any effects due to the ball rotating and changes in lighting (reflections etc). Distortions of the spherical shape of the ball occur due to camera lens radial distortion and the changing viewing angle between the camera and the ball, as it moves. Viewing the ball from below (i.e. the camera's horizontal viewing plane is lower than the ball) will produce a ellipsoidal shape, instead of a sphere. Equally, viewing from above will cause a similar distortion (i.e. the camera's horizontal viewing plane is above that of the ball). (Any change in viewpoint (under an affine transform) about the normal viewing-plane will cause a circle to be perceived as an ellipse.) These effects, along with imperfect extraction causing noise around the perimeter of the basketball, are visible in the extracted images - producing variations in the shape contour between consecutive binary images. Figure 3.8 demonstrates this, where here the binary images have been scaled and centralised (with respect to the ball) to ease visualisation.

A small set of low order Zernike velocity moments was then run on the three sequences. The results were analysed in terms of the corresponding mean μ, standard deviation σ and $\sigma/\mu\,\%$ (coefficient of variation), indicating the percentage spread of the moment values. Table 3.5 shows a selection of these results, demonstrating moments that are tightly clustered (A_{2000}, A_{4200} and A_{3100}). These moments are describing the ball shape itself, its structure, which is highly correlated throughout the three image sequences, although still perturbed by camera distortion, perimeter noise etc. The remaining moments separate the sequences producing $\sigma/\mu\,\% > 100$ (A_{2022}, A_{4220} and A_{3110}). These are describing the structure of the ball, along with the corresponding velocity information - their motion.

This simple experiment has demonstrated the use of both exploiting temporal correlation, and the inclusion of motion in the shape descriptor. The use of temporal correlation has diluted the effect of noise within the image sequence, while

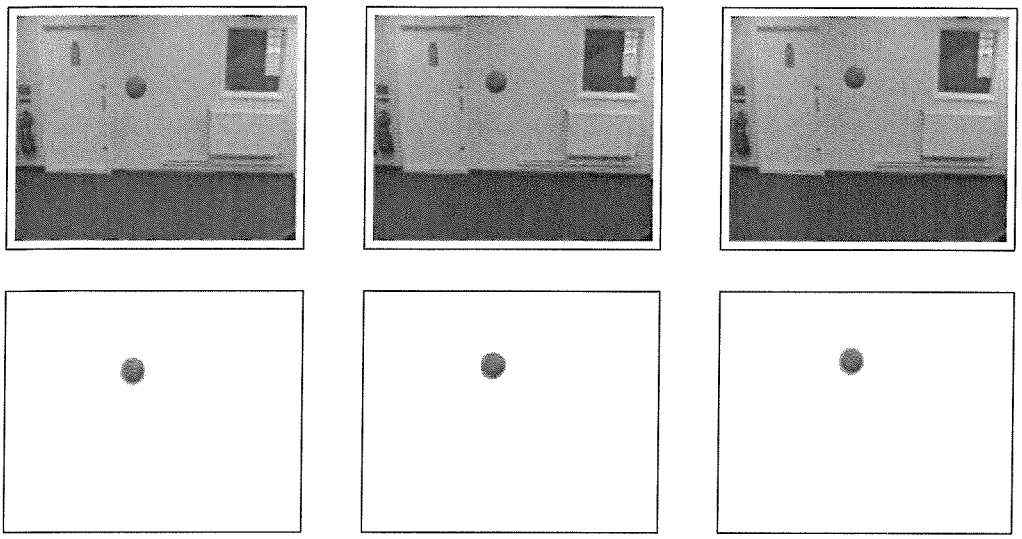


Figure 3.7: Example consecutive images from sequence 3, along with their extracted versions.

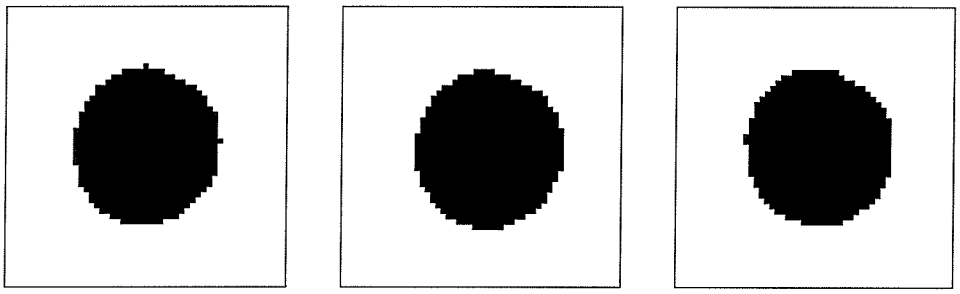


Figure 3.8: The extracted basketball from sequence 3, showing varying shape contours between images.

motion information has enabled the separation of the three image sequences of a moving basketball - producing moments that can describe both shape and motion. The variation in shape contour between consecutive images has been encoded along with the velocity information for each pair of images. In this way the final description holds not only the average velocity information but also a result from the correlation of the different contours and their corresponding velocity. In the simple case of the moving ball, this information may not be of interest. However, if there are larger changes in shape between consecutive images, along with a variation in the velocity component (within the sequence), then this information becomes potentially more interesting. The next chapter on human gait analysis aims to exploit these properties, using the velocity moments to describe a temporal image sequence of a shape, which (as the sequence progresses) alters in both composition and motion.

Sequence	$\overline{A_{2000}}$	$\overline{A_{4200}}$	$\overline{A_{3100}}$	$\overline{A_{2022}}$	$\overline{A_{4220}}$	$\overline{A_{3110}}$
1	1.899	0.129	0.091	6.035	0.018	0.003
2	1.892	0.117	0.065	926.831	1.232	0.018
3	1.899	0.074	0.086	2819.414	11.118	0.138
μ	1.897	0.029	0.081	1250.760	4.123	0.053
σ	0.004	0.107	0.014	1434.390	6.088	0.074
$\sigma/\mu\%$	0.215	26.803	17.277	114.681	147.665	140.369

Table 3.5: Moment clustering results for the three sequences of a bouncing basketball - demonstrating both tight and loose clustering.

3.8 Discussion

It is usually prudent to compare a new technique with an existing equivalent technique, to aid its characterisation. The performance of the Cartesian velocity moments has already been compared with an averaged Hu [28] invariant set, Section 3.2.3. This was performed to help illustrate the advantages of exploiting temporal correlation to overcome problems including image noise. The analysis also illustrated some possible disadvantages of using non-linear combinations of Cartesian moments. Further, the possible advantages of including motion into the descriptor have been addressed, as applied to real-world imagery in Section 3.7. In view of this, to further illustrate and characterise the velocity moments under differing conditions, a set of classification and performance tests are proposed. Firstly, both implementations are analysed with respect to classification problems on a series of human gait databases, Chapter 4. Secondly, the characterisation of the Zernike velocity moments technique under the conditions of image noise, occlusion, image resolution degradation and time-lapse imagery is investigated. This is done to provide the user with an insight into their behaviour under these conditions and is detailed in Chapter 6. Depending on the velocity moment chosen, different characteristics (or information) from an image sequence's distribution will be described. Ultimately there is an infinite number of moments that could be computed for any given image sequence, each of which will invariably behave differently under varying conditions, such as occlusion or image noise. However, the general effect of these conditions should remain constant. For example, the low order Zernike velocity moments, A_{0000} and A_{1100} are set to known values by the mapping process, thus will not be affected by occlusion, unless the object is totally occluded, i.e. there is nothing present in the image. Therefore, including A_{00**} and A_{11**} in any performance tests would bias the results. In general, low order moments describe gross shape information, such as mass or spread in each axis. Higher order moments describe the high frequency components of the image, or fine detail. For any given application, a range of both orders (high and low) is likely to be used. However,

characterising all the velocity moments is not viable, due to the possible list being infinite. The characteristics of the low order Zernike velocity moments have already been examined under the problem of rotation, Section 3.3.3. Previous work has studied the effects of image noise [80], while moments are known to perform poorly under occlusion, due to the description being a single image area measure. If you remove part of the shape, understandably the result will alter. For these reasons, it was decided to observe the effects of varying conditions (i.e. occlusion, noise etc) using the velocity moments that proved useful for a particular application. The chosen application is the use of the Zernike velocity moments for the description and classification of moving shapes, as applied to human gait.

3.9 Conclusions

In this chapter we have presented a new framework which enables the extension of traditional moment theory to include the analysis of image sequences. Previously moment analysis of an image sequence would concentrate on each separate image within the sequence. Here, however, the analysis enables the sequence to be described collectively, while also including shape-motion information. This results in a framework that allows the statistical analysis of both rigid and non-rigid moving shapes, and is next applied to describing walking people.

Chapter 4

Application to human gait

This chapter describes the ideas behind human gait classification, and details previous work in this relatively young field of research. It continues by outlining the approach taken to enable the analysis of gait via the new velocity moments. The chapter then concludes with descriptions of the feature selection and classification techniques to be used in the next chapter, where the results of applying the velocity moments to seven different gait databases is presented.

4.1 Introduction

Gait is defined as the 'manner of walking or forward motion' [81]. It is primarily determined by muscular and skeletal structure. In this way a person's gait can be perceived as being individual, and comprised of hundreds of components. (However, within-subject variations may also exist depending on a person's mood, posture, etc.) Human gait components range from a subject's thigh rotation patterns and leg swings, to their 'bobbing' or vertical motion. To help understand the different parts of a person's gait, we must first explain a gait cycle, as defined by Murray [55]. A gait cycle is the time interval between successive foot-to-floor contact or heel strikes for the same foot, Figure 4.1. Referring to Figure 4.1 (and the right foot), the gait cycle begins at the heel strike, the first part of the *stance* phase. As the ankle flexes the foot is placed flat on the ground, as the subject's body weight is transferred onto it. As the other leg (shown in black in Figure 4.1) swings through to the front, the heel of the first lifts. The knee of the supporting leg flexes to allow the shift in body weight to the other leg. The lifted leg which is behind, lifts further to clear the ground and the *stance* phase ends as the toe (of the lifted leg, shown in white in Figure 4.1) leaves the floor. Next the *swing* phase commences as the toes of the lifted leg leave the floor, this means that the body weight is transferred to the other foot. The swinging leg moves forward to strike the ground in front of the other foot and thus begins the next gait cycle. Two further characteristics of the gait cycle are the *stride* and *step* length. The *stride* length in defined as the

linear distance in the plane of progression between successive points of foot-to-floor contact for the same foot, while the *step* length is the distance between successive contact points of opposite feet.

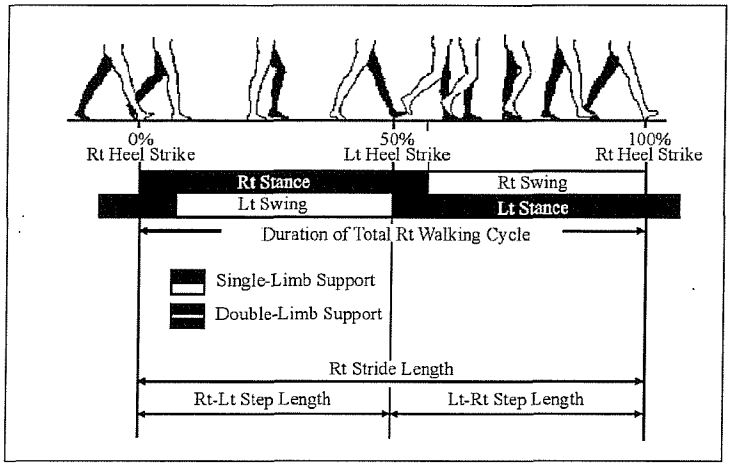


Figure 4.1: Relationships between different gait cycle components.

4.2 Previous work in human gait recognition

Recognition of a person digitally, via biometrics is a method widely used today. These methods aim to capture information about a person's physical characteristics or personal traits. Current methods include fingerprints, iris and retinal scans, hand and face geometry, speech and even key stroke dynamics. Humans use biometrics to identify familiar people, by characteristics like hair colour, face structure and height. Here we are considering using gait as a biometric. It has advantages over current methods as it is both non-invasive and difficult for the subject to hide or disguise. One of the earliest documented examples of recognition by gait was Shakespeare who wrote in The Tempest[Act 4 Scene 1]

"High'st Queen of state, Great Juno comes; I know her by her gait"

There are two main approaches in computer vision to gait recognition. The first is model-based where the subject's movement is described by a mathematical model. This approach was used by Niyogi [59] where recognition of a walking subject was detected by looking at an XT-slice (where X is a slice along the x axis through a stacked image sequence and T is time), from which the trajectory through time of the subject is reconstructed using snakes. This information is then used to create a stick model of the subject for recognition. Nash [57] used a simple pendulum model as a basis for searching a scene to locate a moving person using a method called the Velocity Hough Transform (VHT) [56]. Cunado [12] built on this by using the VHT with a double pendulum model to characterise the hip movement of a subject and from this analysed the frequency response of this simple harmonic motion.

This idea was extended to incorporate the lower leg motion, producing a coupled oscillator gait model [84] that was then applied to a database of running and walking gait sequences. Other modelling techniques include using Kalman filters to filter or track the movement of an individual through a sequence of images [1, 66] and then to analyse this movement (although not yet used for recognition purposes). A more recent study analysed silhouettes from varying view points and, through view point calibration, composed a view-invariant set of body parameters that are used for gait classification [36].

An alternative method (and the one which is used here) is to apply a holistic description to the set of images. This approach has been used by Murase [54] where an eigenspace representation enabled efficient image sequence comparison. First silhouettes of the subject are extracted from the sequences. These silhouettes are then projected into the eigenspace where they are compared with a database of previously analysed sequences. Using this system a recognition rate of 100 % was achieved on a database of seven subjects. A similar approach has been used by Huang [31], where optical flow fields are generated which are then processed using both eigenspace transformation and canonical space transformation, to achieve 100 % recognition. Another study also used eigen analysis to characterise gait [4], however, here the technique was applied to self-similarity maps. Whereas, Meyer [49] used optical flow information to help estimate the shape of the person and from this went on to train Hidden Markov models [50] to describe and classify different types of gait. Alternative non-model based gait recognition techniques include the analysis of temporal symmetry [25, 26] or area-based masks to enable direct extraction of specific gait characteristics [19, 20]. One statistical study [46] again used optical flow fields, but feature description was achieved using a method based on low order moments. From this a model free description of instantaneous motion was achieved.

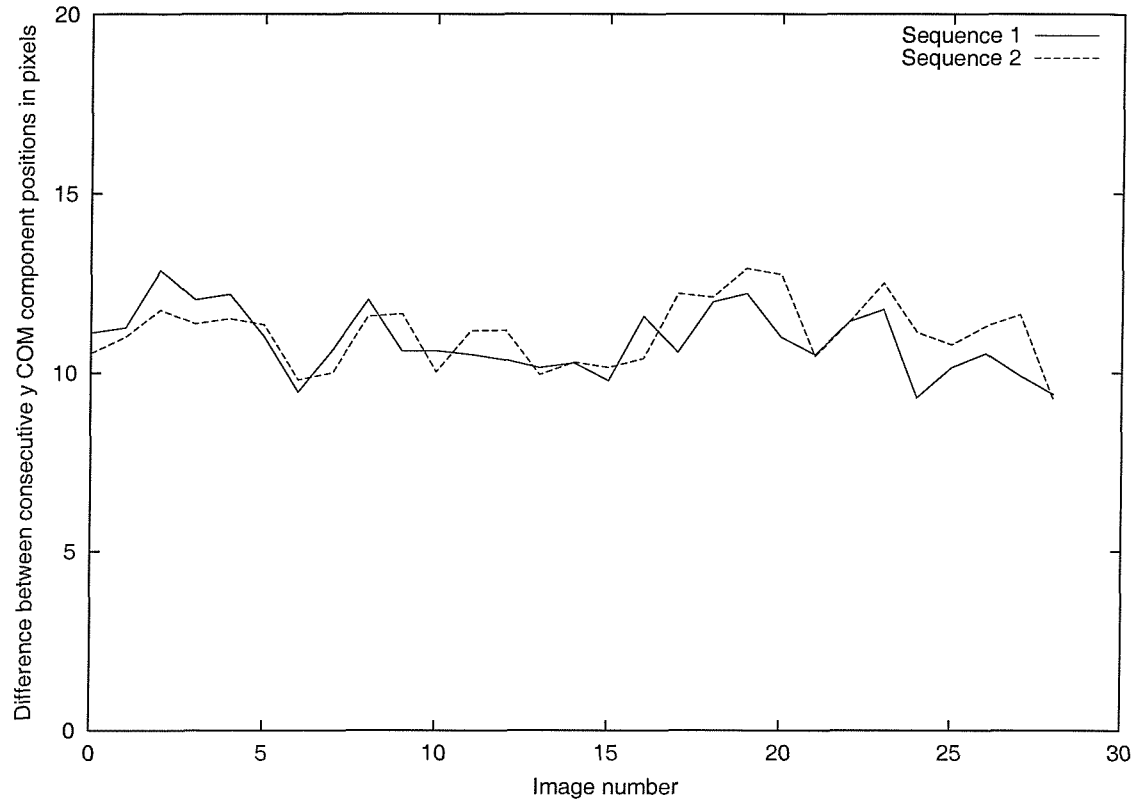
The model based methods are more amenable to re-deployment to alternative camera views, or even different applications. However, the statistical methods, or more specifically the holistic techniques, have improved capability over application problems such as image noise, due to them utilising more subject information i.e. using the subject's complete silhouette, as compared with a model description of just their legs.

4.3 Methodology

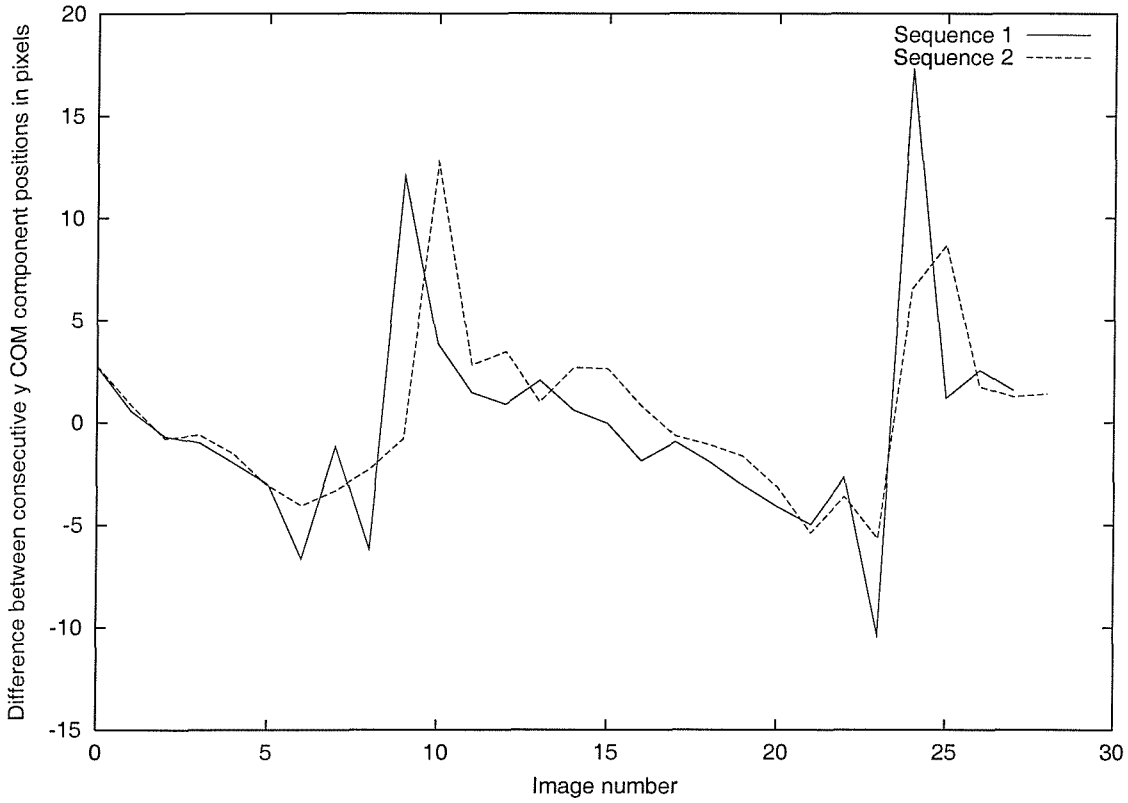
As a person walks, variations in both horizontal and vertical motions exist. Here, we intend to produce descriptions that link both the person's motion and their corresponding shape, in each stage of their gait cycle. The motion information is extracted from the image sequence using their COM (centre of mass). Figure

4.2 shows the corresponding x and y COM plots for two sequences of the same subject, walking for one complete gait cycle (heel strike to heel strike). Figure 4.2b clearly shows the 'bobbing' vertical motion as the subject reaches the minimum leg-width position (both legs together) producing their maximum height in the sequence - around image 10 and again at image 25. It can be seen that variations exist between the two sequences on each plot, especially Figure 4.2a. However, it must be noted that the sampling rate is 25 frames per second, meaning that the first sequence samples may well be missing from the second sequence and *vice versa*. This possible under-sampling is less apparent in the Figure 4.2b where the variations in motion are greater in magnitude. (There may also be between-sequence variations introduced by the extraction technique). These variations in COM are linked to the images themselves, suggesting that using just the x or y COM variations alone would not prove useful (in terms of classification).

Humans perceive gait by observing a person's overall shape and how this moves and changes as they walk. Thus, for human vision both shape and motion are important. Consequently, we duplicate this behaviour in our classification approach by using two different image sets. Firstly a set of binary silhouettes, or spatial templates (STs) for each subject sequence is obtained through removal of the background. Optical flow images or temporal templates (TTs) are then computed using an algorithm based around matching image patches, the results of which are consistent with human psychophysics, [8]. Further, this method of determining optical flow produces results similar to how humans perceive motion. The velocity moments are then calculated for these two image sets, STs and TTs. Due to the periodic nature of gait, analysis is performed on one complete gait cycle (Figure 4.1). In this way, the results of classification are not biased by unbalanced consideration of different parts of the gait cycle. However, a sequence of images containing data from 'heel strike to heel strike' will contain a duplicate image - the first and last image will be identical. The velocity moments use the first image to calculate motion information only, so this duplication will not bias the overall calculation. Those velocity moments suitable for classification are then selected using the single factor ANOVA technique and the Scheffe post-hoc test. Due to the small number of gait sequences (or samples) available per subject, the ANOVA method is only used as a guide as the resulting variance estimates will be inaccurate. This analysis is viable for small databases due to the small number of subjects. However, using these techniques with greater numbers of subjects proves impractical due to the increased number of features needed to separate them. The single-way ANOVA selects features that singularly separate portions of the dataset. Whereas we are actually using multiple features to classify, suggesting an n-way ANOVA may prove more useful for larger datasets, enabling the analysis of the interaction between



(a) x COM (average velocity removed).



(b) y COM.

Figure 4.2: The x and y COM variations for one complete gait cycle (heel strike to heel strike of the same foot).

features for an n dimensional feature or classification space. For larger databases, the single factor ANOVA analysis can be useful in reducing the feature set to those features which *may* prove useful for classification. Ideally, an exhaustive search of all combinations of this reduced feature set can then be performed using the classification rate as the measure of success. Unfortunately, even if the reduced feature set contains just 100 possible moments, solving this problem (in terms of all possible combinations of moments) would take months on a 1 GHz machine (for a database of 200 sequences). Here seven databases are analysed, making the exhaustive approach impractical due to time constraints. Further, the main drive of this research is not feature set selection, a vast research field in itself. For these reasons final selection is achieved by manual intervention, resulting in possible non-optimal results (i.e. an entirely different subset of the reduced moment set may produce an improved or identical classification rate). The selected moments are used to produce a multidimensional feature space for classification, rather than combining them (i.e. non-linear combinations of moments) prior to classification. Combining features in this manner is avoided for the reasons discussed in Section 3.2.3.

Classification of these selected features is possible through a number of different methods, ranging from simple distance metrics to neural networks and support vector machines. Again, classification theory is itself a large field of research. Here we have chosen to use a simple classifier so as to avoid getting trapped in the intricacies of classifier theory. Thus, simple classification (or recognition) of the moment features is achieved using the *k-nearest neighbour* technique ($k = 1$ and $k = 3$) using the leave one out rule with cross validation. Doubtless the overall classification results could be improved by using a more powerful technique.

4.4 Template extraction

To apply this new method to gait, we first need to extract the subject (or feature). Various different methods of extraction were used throughout this work. The differing methods reflect a progression of new extraction techniques and different scenarios. Using different techniques illustrates that the results gained from the velocity moments are not dependent solely on one extraction technique. In all, seven databases were analysed: SOTON (University of Southampton), UCSD (University of California, San Diego), four CMU (Carnegie Mellon University) databases and finally the HiD dataset (Human ID at a distance program - captured at the University of Southampton). However, the differing extraction techniques follow the same overall structure. Simple feature extraction is achieved using templates, first spatial and then temporal. Silhouette data or spatial templates (STs) are produced, which in turn are used to extract the subjects, allowing the calculation of optical flow or

temporal templates (TTs). These then provide suitable data for study using the velocity moments.

In the next sections we describe the different extraction methods employed. However, it must be noted that many alternative model-based and statistical-based extraction techniques exist, eg. [24, 57].

4.4.1 Subject extraction 1 - Background subtraction

This simple subtraction method was applied to the SOTON database. The complete method of template extraction is shown in Figure 4.3. A background image is derived by application of a temporal-mode filter to the image sequence. By selective subtraction and region growing a subject can be extracted, which then can be used to compute optical flow (described in Section 4.4.4). For the extraction of spatial templates (STs), the subject is first isolated using background subtraction. However the difference image thus produced is prone to noise, i.e. speckle noise in the background scene, or holes appearing in the subject's silhouette. To remove this noise the image is region grown. Merging of pixels is determined by evaluating a homogeneity criterion. The region growing algorithm is a variant of the basic split and merge method. However instead of using a hierarchical data structure, which is then searched to determine areas to be merged, merging is achieved by looking at the pixel distribution, a variation of the algorithm proposed by Dubuisson and Jain [16]. Finally to produce a binary spatial template that is suitable for use in extraction, the region grown image is thresholded. The final silhouette can then be windowed using the subject's average velocity. This velocity value is then stored for later use. Logically ANDing the silhouette with the original grey-scale image allows subject extraction, as shown in Figure 4.3.

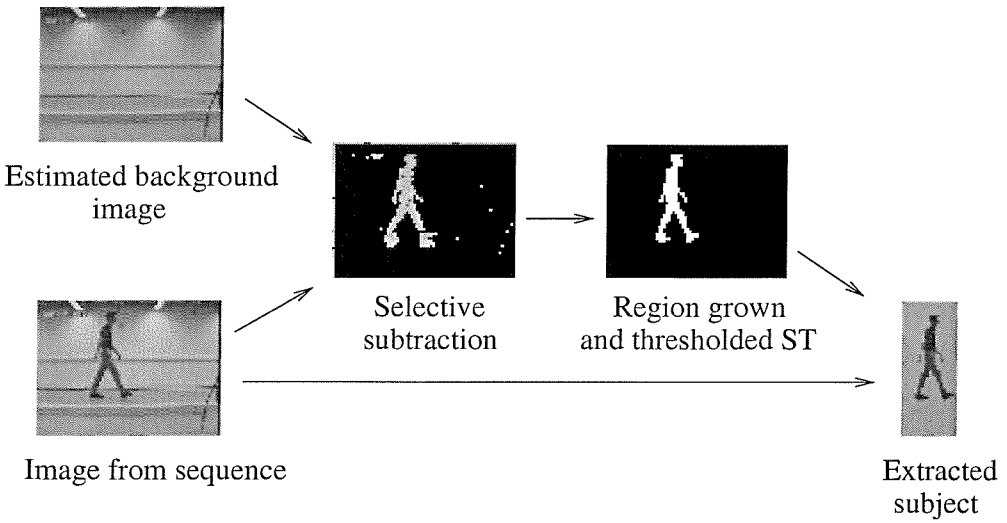


Figure 4.3: Producing the spatial templates (STs).

4.4.2 Subject extraction 2 - Statistical scene analysis

For the UCSD database a statistical based subject-extraction method [33] was used to produce a small database of silhouettes. The extraction method analyses the statistics of the sequence and uses both luminance values and edge information to determine the background and foreground objects. First, a background model is constructed which consists of time averaged mean and variance images. A modified background subtraction technique based on these mean and variance images produces a set of confidence maps. These two steps are repeated for both the luminance and edge information. The results of which are combined to produce a final set of confidence maps. These in turn allow the separation of the foreground (moving) objects. Appendix B contains a more detailed explanation and example silhouettes.

4.4.3 Subject extraction 3 - Chroma-keying

Due to the nature of both the capture and colour data of the HiD database, the use of a colour specific extraction was possible. Blue screening or more generally chroma-keying is the process of filming an object, or subject in front of an evenly lit, bright pure coloured backdrop. Object or subject extraction (through backdrop removal) is easily achieved, allowing an alternative background colour or scene to be used in its place. Any pure colour can be used as the backdrop, although the choice is restricted by the colours of interest on the subject. Here a bright green was used, mainly as it is an unlikely colour for the subjects to wear. Also, video cameras are usually more sensitive in the green channel, and often have the best resolution and detail in that channel (due to having twice as many green pixels as red and blue, an attempt at matching human-vision colour sensitivity).

Chroma-key laboratory lighting

To maintain a uniform chroma colour backdrop (thus improving extraction) two lighting schemes are needed, one for the subject and the second for the backdrop. To enable an evenly lit backdrop and maintain an illuminated subject while reducing the effects of shadows, the lighting schemes are separated. Powerful flood lighting is used to light the backdrop, preferably from directly above and from the sides. The subject is then placed in front of this lighting, while a further set of lower power diffuse lights are used to light the subject directly, refer to Figure 4.4.

Chroma-key software extraction

The chroma-key process is based on the luminance key. In a luminance key, everything in the image over or under a set brightness level, is 'keyed' out and replaced by either another image, or a colour from a colour generator. Here we key out the (bright-green) colour to remove the backdrop and the floor. An absolute error range is added around the selected colour to allow for lighting variations due to

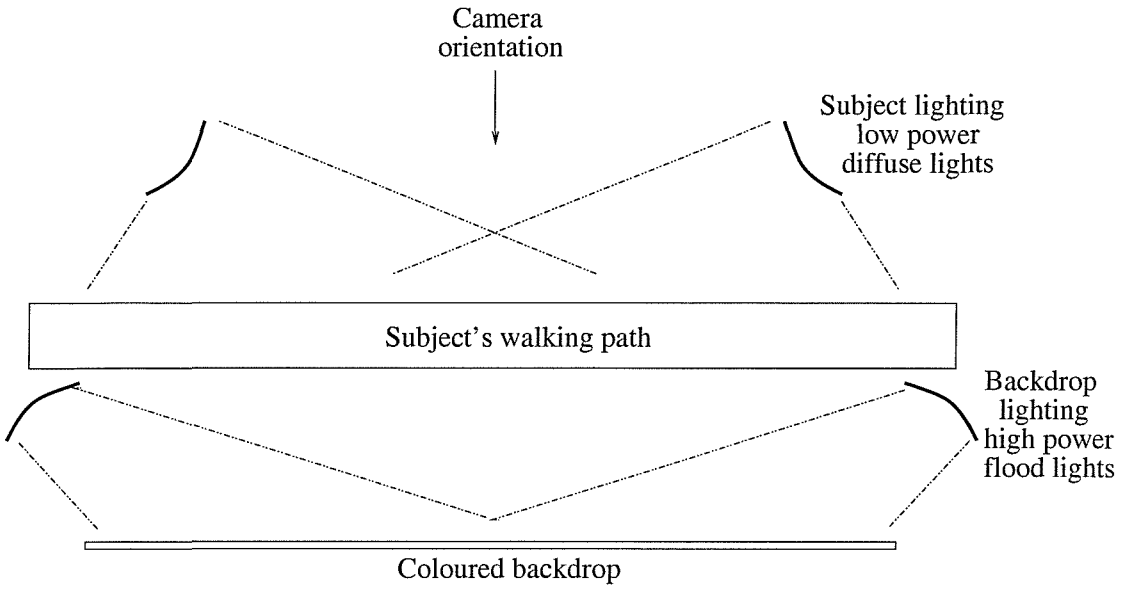


Figure 4.4: Laboratory lighting arrangement, enabling the separation of the two lighting schemes.

the large backdrop. Noise due to isolated pixels within the backdrop area can be removed by a simple shrink and expand operation. This process will leave large objects relatively untouched. Figure 4.5 shows an example source and keyed image from a sequence. It can be seen that part of the ceiling (a mirrored ceiling light reflector) has been removed due to reflections from the backdrop colour.

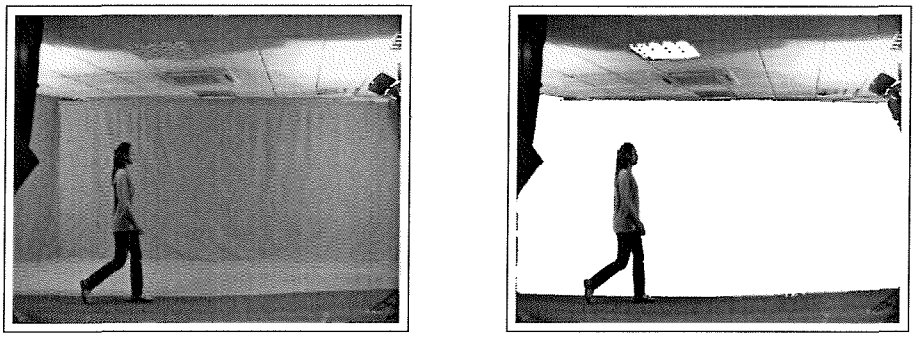


Figure 4.5: Example original image and chroma-keyed result.

Chroma-key spatial templates

By processing the image with a second stage chroma-key the (darker green) floor is removed. Using simple thresholding on the result produces a binary silhouette, Figure 4.6. This result can be further improved, first by cropping the image to remove superfluous data. A simple connected components algorithm along with filtering by size removes all but the largest object (the subject), Figure 4.6. Finally, any holes left by the two passes of the chroma-key extraction can be filled by a simple expanding and shrinking process.

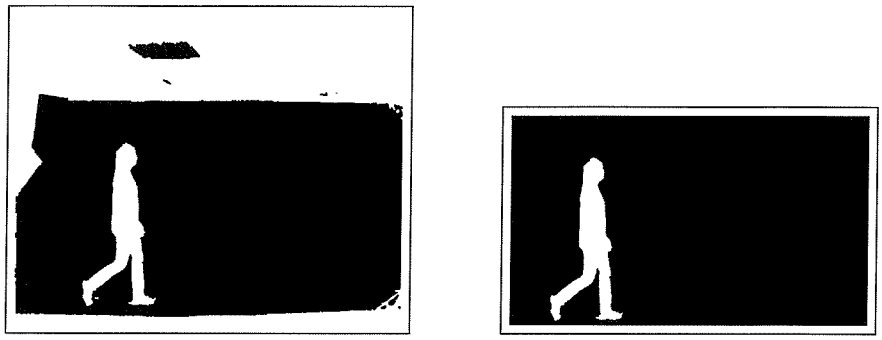


Figure 4.6: Example silhouette and final cropped ST.

4.4.4 Dense optical flow fields

Using the STs as a mask, the subject can be extracted from the original intensity image - the result can be used to produce an optical flow description of the subjects motion. Here we use Dense Optical Flow Fields (DOFF) [8], generated by minimising the sum of absolute differences between image patches. The algorithm relies on the assumption that the optical flow is due locally to a first approximation to fronto-parallel translation of a Lambertian surface. To this end the images are first filtered to remove the effects of shadows, reflections and changes in lighting. This is achieved, first by taking the logarithm of brightness and then by filtering using a Laplacian of Gaussian function, or band pass filter. The corresponding patch for each pixel is compared with a finite number of shifted versions of the original image n, the results of which form a voting space. The shifted patch producing the highest correlation with the $(n+1)$ image's equivalent patch, determines the motion for that pixel. Therefore, two consecutive images produce one optical flow image, Figure 4.7. The images are then reversed and the process is repeated. Upon summing the results of the two passes, correct results will cancel, only these pixels are retained. This helps to reduce any possible incorrect estimates. In this way DOFF images are produced for both the x and y directions, the magnitudes of which are then added to produce the final Temporal Template (TT), shown in Figure 4.8. White pixels correspond to no detected movement, light-grey pixels represent high amounts of motion, conversely darker areas signify low amounts of motion. It has been displayed here in this way to aid visualisation. The areas corresponding to the subject's legs and hands are lighter in colour than the torso, indicating (as expected) greater movement. The magnitudes of $x + y$ flow ($\mid x + y \mid$) are used as Huang [30] showed that these had improved descriptive capabilities than just x flow or y flow alone.

The algorithm assumes that only differences due to motion are detectable between consecutive images. As a result it is limited by the pixel size relative to the motion being described. I.e. if the motion is poorly represented due to the low image resolution, then it cannot be described effectively or if the motion is too large

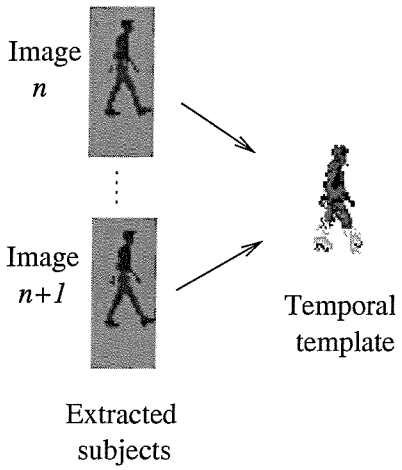


Figure 4.7: Producing the temporal templates.

Figure 4.8: Example consecutive temporal templates.

then it may be lost. As a result the chosen patch and shift sizes are dependent on the image size, the object size (producing the motion) and the motion itself.

4.5 One way ANOVA - Analysis of variance

Analysis of variance (ANOVA) is a general method for studying sampled-data relationships [9, 10]. The method enables the difference between two or more sample means to be analysed, achieved by subdividing the total sum of squares. One way ANOVA is the simplest case. The purpose is to test for significant differences between class means, and this is done by analysing the variances. Incidentally, if we are only comparing two different means then the method is the same as the $t-test$ for independent samples. The basis of ANOVA is the partitioning of sums of squares into between-class (SS_b) and within-class (SS_w). It enables all classes to be compared with each other simultaneously rather than individually; it assumes that the samples are normally distributed. The one way analysis is calculated in three steps, first the sum of squares for all samples, then the within class and between class cases. For each stage the degrees of freedom df are also determined, where df is the number of independent 'pieces of information' that go into the estimate of a parameter. These calculations are used via the Fisher statistic to analyse the

null hypothesis. The null hypothesis states that there are no differences between means of different classes, suggesting that the variance of the within-class samples should be identical to that of the between-class samples (resulting in no between-class discrimination capability). It must however be noted that small sample sets will produce random fluctuations due to the assumption of a normal distribution. If d_{ij} is the sample for the i^{th} class and j^{th} data point then the total sum of squares is defined as:

$$SS_t = \sum_{i=1}^{S}\sum_{j=1}^{D}(d_{ij} - GM)^2 \tag{4.1}$$

with degrees of freedom:

$$df_t = (S\ D) - 1 \tag{4.2}$$

where D is the number of data points (assuming equal numbers of data points in each class) and S is the number of classes and GM is the grand mean:

$$GM = \frac{1}{(S\ D)}\sum_{i=1}^{S}\sum_{j=1}^{D} d_{ij} \tag{4.3}$$

The second stage determines the sum of squares for the within class case, defined as:

$$SS_w = \sum_{i=1}^{S}\sum_{j=1}^{D}(d_{ij} - M_i)^2 \tag{4.4}$$

where M_i is the i^{th} class mean determined by:

$$M_i = \frac{1}{D}\sum_{j=1}^{D} d_{ij} \tag{4.5}$$

and the within class df is:

$$df_w = S(D-1) \tag{4.6}$$

The sum of squares for the between class case is:

$$SS_b = \sum_{i=1}^{S} D\ (M_i - GM)^2 \tag{4.7}$$

with the corresponding df of:

$$df_b = S - 1 \tag{4.8}$$

Defining the total degrees of freedom df_t and the total sum of squares SS_t as:

$$df_t = df_b + df_w \tag{4.9}$$

$$SS_t = SS_b + SS_w \tag{4.10}$$

Finally if MSS_b is the mean square deviations (or variances) for the between class case, and MSS_w is the reciprocal for the within class case then:

$$MSS_b = \frac{SS_b}{df_b} \quad ; \quad MSS_w = \frac{SS_w}{df_w} \tag{4.11}$$

It is now possible to evaluate the null hypothesis using the Fisher or F statistic, defined as:

$$F = \frac{MSS_b}{MSS_w} \tag{4.12}$$

If $F >> 1$ then it is likely that differences between class means exist. These results are then tested for statistical significance or P-value, where the P-value is the probability that a variate would assume a value greater than or equal to the value observed strictly by chance. If the P-value is small (eg. $P < 0.01$ or $P < 1\,\%$) then this implies that the means differ by more than would be expected by chance alone. By setting a limit on the P-value, (i.e. $1\,\%$) a critical F value can be determined. The critical value F_{crit} is determined (via standard lookup tables) through the between-class (df_b) and within-class (df_w) df values. Values of F greater than the critical value denote the rejection of the null hypothesis, which prompts further investigation into the nature of the differences of the class means. In this way ANOVA can be used to prune a list of features. Figure 4.9 shows example F_{crit} values for a low df distribution (i.e. a small dataset). As df increases (i.e. the dataset size increases) the F distribution will become 'tighter' and more peaked in appearance, while the peak will shift away from the x axis towards $F = 1$.

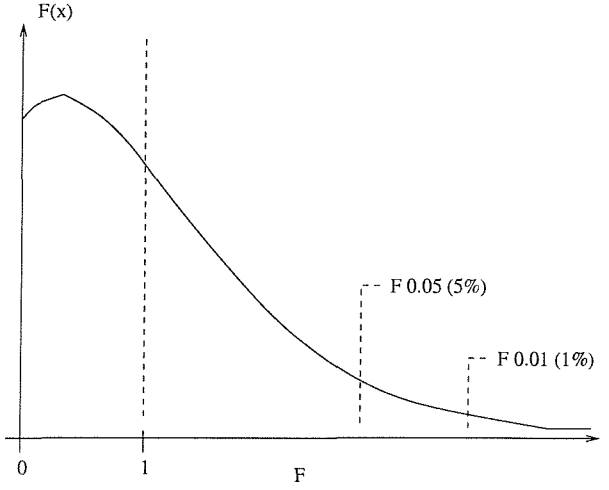


Figure 4.9: Example F distribution (for low df) showing possible $5\,\%$ and $1\,\%$ intervals.

The F value gives a reliable test for the null hypothesis, but it cannot indicate which of the means is responsible for a significantly low probability. To investigate

the cause of rejection of the null hypothesis post-hoc or multiple comparison tests can be used. These examine or compare more than one pair of means simultaneously. Here we use the Scheffe post-hoc test. This tests all pairs for differences between means and all possible combinations of means. The test statistic F_S is:

$$F_S = \frac{(\overline{M_i} - \overline{M_j})^2}{MSS_w \left(\frac{1}{n_i} + \frac{1}{n_j}\right) df_b} \tag{4.13}$$

where i and j are the classes being compared and n_i and M_i are the number of samples and mean of class i, respectively. If the number of samples (data points) is the same for all classes then $n_i = n_j = D$. The test statistic is calculated for each pair of means and the null hypothesis is again rejected if F_S is greater than the critical value F_{crit}, as previously defined for the original ANOVA analysis. This Scheffe post hoc test is known to be conservative which helps to compensate for spurious significant results that occur with multiple comparisons [10]. The test gives a measure of the difference between all means for all combinations of means.

In terms of classification, large F statistic values do not necessarily indicate useful features. They only indicate a well spread feature space, which for a large dataset is a positive attribute. It suggests that the feature has scope or 'room' for more classes to be added to the dataset. Equally, features with smaller F values (but greater than the critical values F_{crit}) may separate a portion of the dataset, which was previously 'confused' with another portion. Adding this new feature, increasing the feature space dimensions, may prove beneficial. In this manner, features which appear 'less good' (i.e. lower F statistic values than alternative features) may, in fact, prove useful in terms of classification.

4.6 Classification

Classification is the method by which a set of measurements are attributed a class label. There are many approaches to this problem including Artificial Neural Networks and simple distance metrics. Classification can also be achieved by simplifying the data set to just those items that contribute to the classification process, by methods like Canonical Analysis or Principle Components Analysis. Here we are concentrating on using the simple k-nearest neighbour classifier (k-nn), by measuring the distance between feature points.

The k-nn classifier associates the sample or feature point with the label of the majority (or mode) of its nearest k-neighbours, if $k = 1$ then the sample is grouped with the class of its nearest neighbour. To determine which neighbours are closest

a distance metric is used. There are many different distance metrics (i.e. Bhattacharyya, Matusita, L_1 norm), here we are using the Euclidean distance d, measured between H feature points, defined as:

$$d = \sqrt{\sum_{i=1}^{H}(s_i - k_i)^2} \tag{4.14}$$

where s is the current sample of interest and k is a known class member (or training point). The dimensionality of the feature space is determined by H, making it easy to add multiple features. Equation 4.14 is also referred to as the L_2 norm. By computing all the distances from the current sample of interest to all other samples, and then arranging them in order of size, the nearest neighbours are determined. By then applying the leave one out rule with cross-validation, all samples are tested against each other. The leave one out rule refers to the method of retaining one sample for test data and using the remaining samples to form the training data. By repeating this for all the samples cross-validation is achieved. These procedures are well established in pattern recognition and allow for appropriate comparison between other classification methods. It must be noted that if between-class feature values vary in their order of magnitude, then prior to classification they may need to be normalised using their maximum values, so as to remove any possible bias.

4.7 Moment order

In general, low order moments describe the gross image information, including image mass and pixel spread. Higher order moments describe the high detail, or high frequency components of the image, (equivalent to the high order harmonics of a Fourier transformed signal). However, the higher order moments are more prone to the effects of noise. This is due to the calculations themselves and also due to the low power of the information being described (i.e. high frequency detail). This presents a problem - at what point are high order moments too noisy to be usable in terms of discriminatory capability? - one approach, used by Khotanzad [37], applied image reconstruction to this problem. By reconstructing the original image, and then studying the pixel error between the original image and the reconstructed one, as the moment order increased - an optimum order can be established. Khotanzad used small (64×64) images of binary typed characters and silhouettes of the 'Great Lakes' for testing. It was found that to achieve a maximum of a 10 % pixel difference (between the original and reconstructed image), order 12 (47 moments) were needed for the typed characters. The Lake dataset required order 8 (23 moments) to produce the same accuracy. Further to this, Teague [79] used up to order 18 (of Legendre moments) for reconstruction of a similar set of typed characters (21×21),

comparing the results visually. Pawlak [62] also studied the effects on reconstruction of increasing the order of orthogonal Legendre moments, demonstrated for a simple 21×21 binary image of a cross, in the presence of Gaussian noise. Again, the pixel error between the reconstructed and original images reduced as the moment order increased. However, a minimum was reached at around order 10, after which the error then increased. These studies have demonstrated the need for relatively high order moments to efficiently describe small images ($\leq (64 \times 64)$). These results suggest the need for different maximum moment orders is dependent upon the application - different applications will produce different image sizes and content. Thus, the maximum order for efficient description can be described as being *data driven*. Here we are studying images of people, where the height of the subjects within the images range from $\simeq$ 90 to 140 pixels (as compared with the 64 pixels of Khotanzad's typed characters). The image sequences are also likely to be rich in high frequency information, indicating a need to study both low and high order moments. However, it must be noted that here we are only interested in separating the classes within a dataset, and not with reconstructing images.

4.8 Conclusions

This chapter has detailed the ideas and methods behind current methods of human gait classification. We have then proposed a method based on human perception of gait, with the aim of utilising both shape and motion information. Simple pre-processing, or subject-extraction methods have been described, providing features which are suitable for analysis by the velocity moments. Finally, methods of feature selection, or more specifically feature list reduction have been described, along with a simple classification method. All these provide the basis for the the next chapter, the analysis of seven human gait databases using both the Cartesian and Zernike velocity moments.

Chapter 5

Database results

5.1 Introduction

This chapter details the analysis of the velocity moments (both Cartesian and Zernike) as applied to seven different gait databases. The subject extraction techniques discussed in the previous chapter are used to provide features suitable for analysis by the velocity moments, while the analysis of the features themselves uses the statistical techniques discussed. A description of each database is included and the results are presented in terms of classification analyses, detailing between-class separation and within-class clustering through the use of the ANOVA technique and the Scheffe post-hoc test. These results are then displayed using scatter diagrams to aid data visualisation. Unless otherwise stated all velocity moment values are normalised, as per Equations 3.7 and 3.16.

5.2 Cartesian velocity moments

We begin by applying the Cartesian velocity moments (Equation 3.2) to the problem of gait classification. The Cartesian velocity moments are designed to capture both spatial and motion information from a sequence of images through use of the non-orthogonal Cartesian centralised moments.

5.2.1 SOTON database

The first of the databases studied is part of the SOTON database, captured at the University of Southampton. It consists of 4 subjects with 4 sequences of each subject. Figure 5.1 shows an example image from the SOTON database. In each sequence, the subject is walking indoors (in a laboratory environment) normal to the direction of the camera, the direction of travel is left to right for two of the four sequences and right to left for the remainder. Each image sequence has been cropped so that it consists of one complete gait cycle, heel strike to heel strike of the same foot. Using the technique detailed in Section 4.4.1, a small database of STs was produced from the SOTON database of image sequences. The resulting STs

were then windowed using the subject's average velocity, the values of which form part of the database. Next, using the STs and the technique described in Section 4.4.4 a small database of TTs was produced. Figure 5.2 shows example templates from the database.

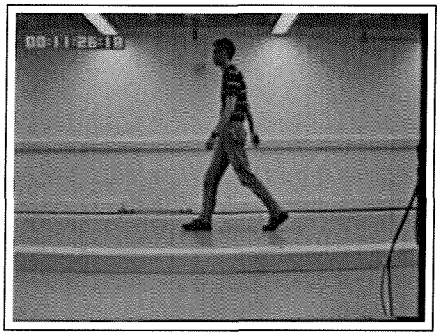


Figure 5.1: Example image from the SOTON database.

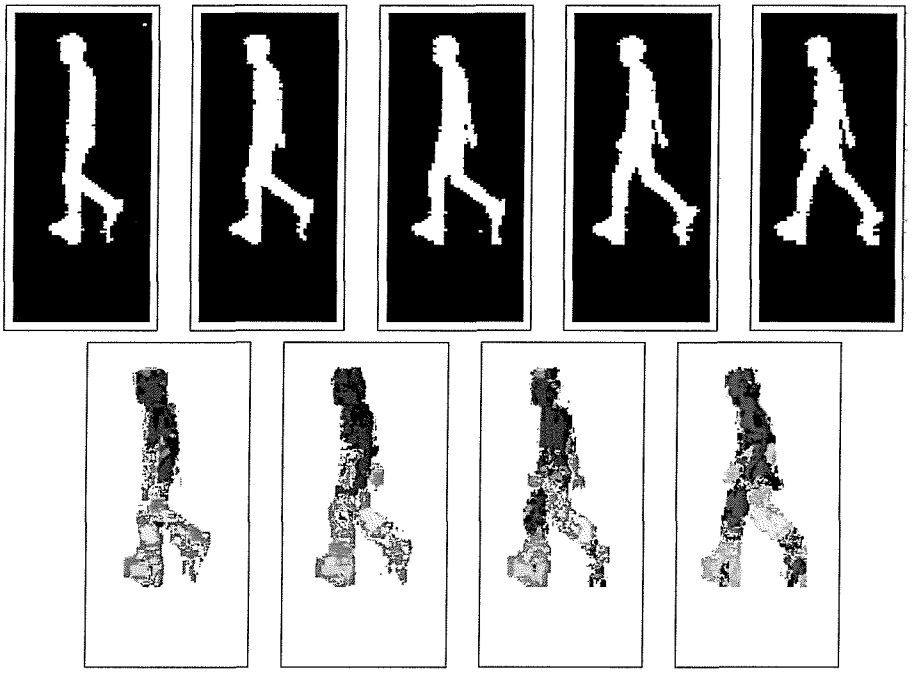


Figure 5.2: Example windowed STs (top) and TTs (bottom) from the SOTON database.

A complete set of Cartesian velocity moments up to $p = q = \mu = 4; \gamma = 0$ (a total of 125 moments) was then calculated on each template set - STs and TTs. In order to help reduce the size of the computational problem, it was decided to only study the x direction velocity information, hence $\gamma = 0$. The average magnitude of velocity (from the windowing operation) for each subject is combined with the actual COM calculation on each image. This combination is then placed into the motion half of the velocity moment calculation (Equation 3.4). The COMs are calculated and added (or removed) to allow for differences which may be present between the average velocity and the actual velocity between successive images.

Velocity moments less likely to be suitable for classification were removed from the calculated list using the single factor ANOVA technique (described in Section 4.5). These were moments producing F statistic values below the 1 % confidence level. Table 5.1 displays the number of moments (in terms of percentages of the total of 125) that showed promising properties in terms of subject clustering and separation, determined via the F statistic. A large majority of these include velocity information. Here $F >> F_{crit}$ illustrates those with $F > 30$ (an arbitrarily large number ensuring $F >> F_{crit}(1\,\%)$ confidence level). Next, using the Scheffe

Template	$F > F_{crit}$	$F >> F_{crit}$
ST	51 % (40 %)	9 % (4 %)
TT	41 % (30 %)	10 % (4 %)

Table 5.1: Percentages of total moments (for the SOTON database) calculated which show promising properties for classification, as determined by the 1 % F_{crit} value of 5.95 - those including velocity information are shown in brackets.

post-hoc test, the differences between sample means of the remaining moments were analysed. Using these results as a guide, manual moment selection from the available set was achieved. Those moments describing both structural and velocity information were favoured, as these descriptors imply a within-subject (or class) correlation, and between-subject separation, between the image sequence and the subject's relative motion - thus focusing on classification by body shape and motion. Three velocity moments (vm_{0200}, vm_{2310} and vm_{2320}) were found sufficient to separate the complete database, illustrated in Tables 5.2a,b and c which show the results from the Scheffe post-hoc test for the three selected moments. Here, each entry is the F_S statistic value for the subjects indicated by the row and column labels (with duplicate comparisons removed). By using the three moments, all combinations are well separated with respect to the 5 % confidence level - only one combination (1,2) is not covered by the 1 % confidence level. The resultant F statistic values for each of the moments selected for the STs can be seen in Table 5.3a, in comparison to the critical values shown in Table 5.3b. Table 5.2d shows the Scheffe result for vm_{2300}, comparing these results with those for vm_{2310} (Table 5.2a) illustrates that the inclusion of velocity has improved the separation of the (3,4) combination, while slightly reducing the separation of the remaining combinations.

Finally, classification of these manually selected moments is achieved using the k-nn approach (with $k = 1$ and $k = 3$), as described in Section 4.6. (Prior to classification the moment values are normalised using their maximum values, so as to remove any bias caused by sets of moments with greater values). Plotting the three selected moments shows distinct subject clustering, Figure 5.3a. While Figure 5.3b demonstrates the linear relationship between vm_{2310} and vm_{2320}. Using

the three selected velocity moments for classification produces the results in Table 5.4. All subjects have been successfully separated producing 100 % classification.

	2	3	4
1	3.5	116.0*	67.4*
2		79.1*	40.1*
3			6.6*

(a) vm_{2310}

	2	3	4
1	2.0	63.9*	26.5*
2		43.4*	14.0*
3			8.1*

(b) vm_{2320}

	2	3	4
1	3.6	9.6*	34.5*
2		24.9*	60.5*
3			7.7*

(c) vm_{0200}

	2	3	4
1	4.1	137.7*	109.7*
2		94.3*	71.4*
3			1.7

(d) vm_{2300}

Table 5.2: Scheffe post-hoc results (F_S) for the SOTON STs. A * indicates a value greater than the 1 % F_{crit} value of 5.95.

Moment	**F-value**
vm_{2310}	156.34
vm_{2320}	78.94
vm_{0200}	70.43

(a) STs.

Confidence	F_{crit}
5 %	3.49
1 %	5.95

(b) F_{crit} values.

Moment	**F-value**
vm_{0400}	88.40
vm_{0200}	154.68
vm_{4110}	52.37

(c) TTs.

Table 5.3: F and F_{crit} values for the selected Cartesian velocity moments on the SOTON database.

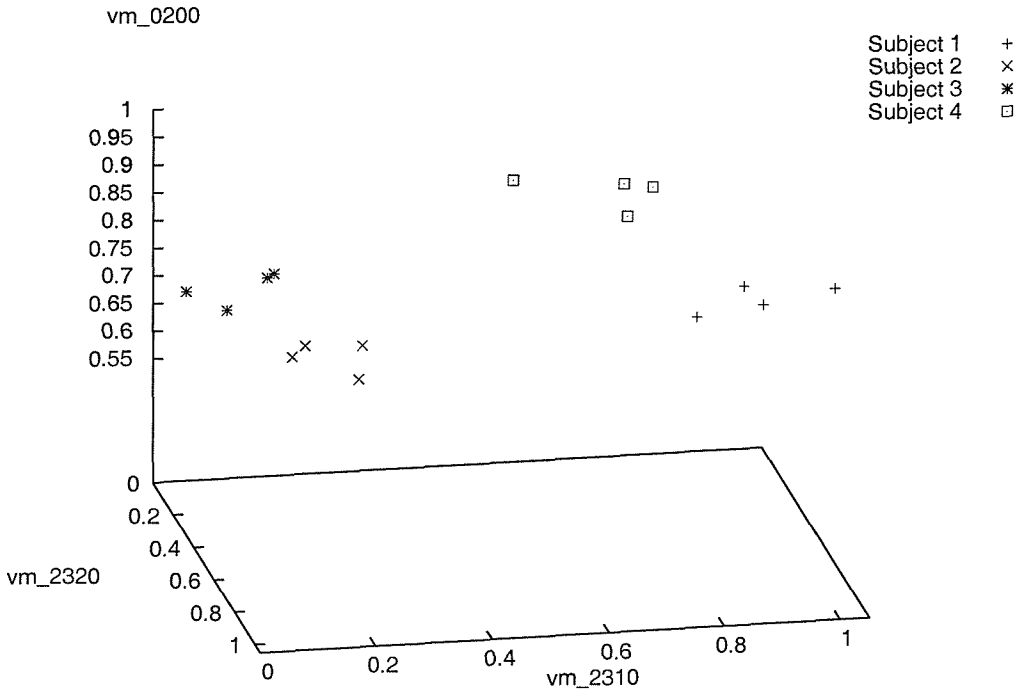
Cartesian velocity moments	**Classification**	
	$k = 1$	$k = 3$
vm_{2310}	93.75 %	87.50 %
vm_{2310}, vm_{2320}	93.75 %	93.75 %
vm_{2310}, vm_{2320}, vm_{0200}	100.00 %	100.00 %

Table 5.4: ST classification results for the SOTON database.

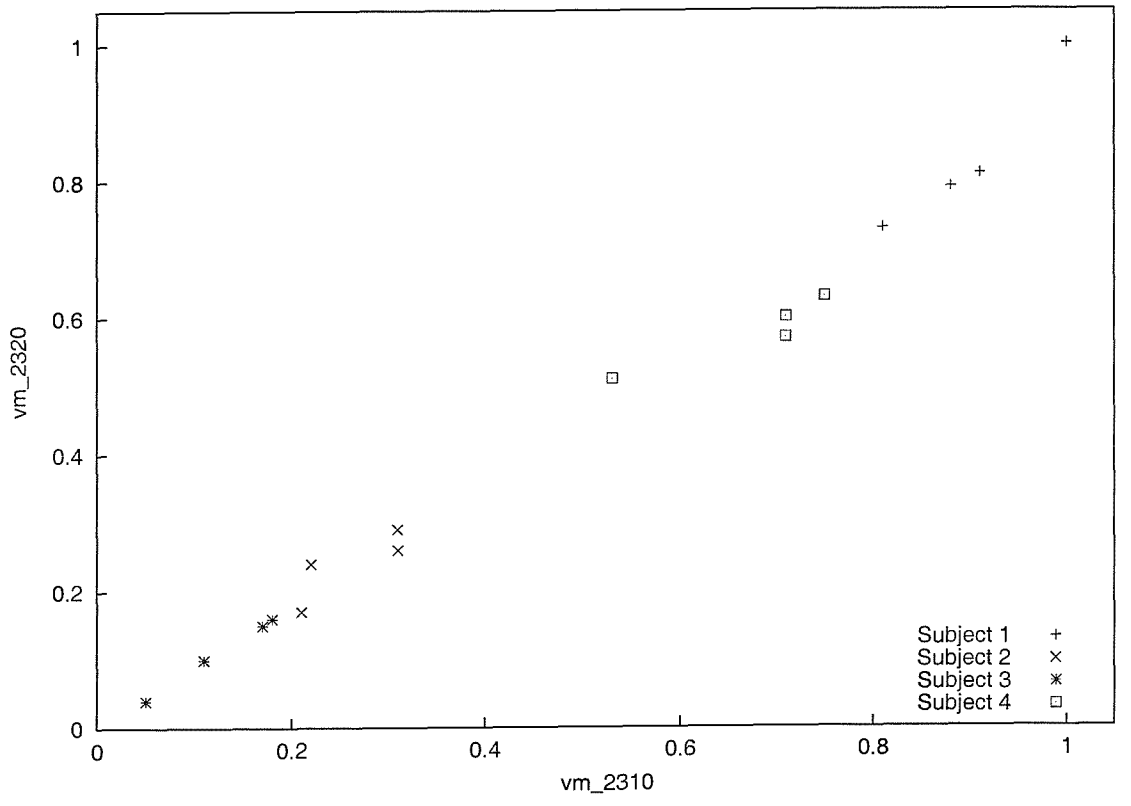
Cartesian velocity moments	**Classification**	
	$k = 1$	$k = 3$
vm_{0400}	87.50 %	81.25 %
vm_{0400}, vm_{4110}	100.00 %	81.25 %
vm_{0200}, vm_{0400}, vm_{4110}	100.00 %	100.00 %

Table 5.5: TT classification results for the SOTON database.

Applying the same techniques 100 % classification is achieved on the TT database, using three velocity moments (vm_{0200}, vm_{0400} and vm_{4110}) as shown in Table 5.5. Plotting these features produces the distinct subject clusters shown in Figure 5.4. Table 5.3c shows that these three velocity moment F statistic values are all far greater then the F_{crit} values shown in Table 5.3b.

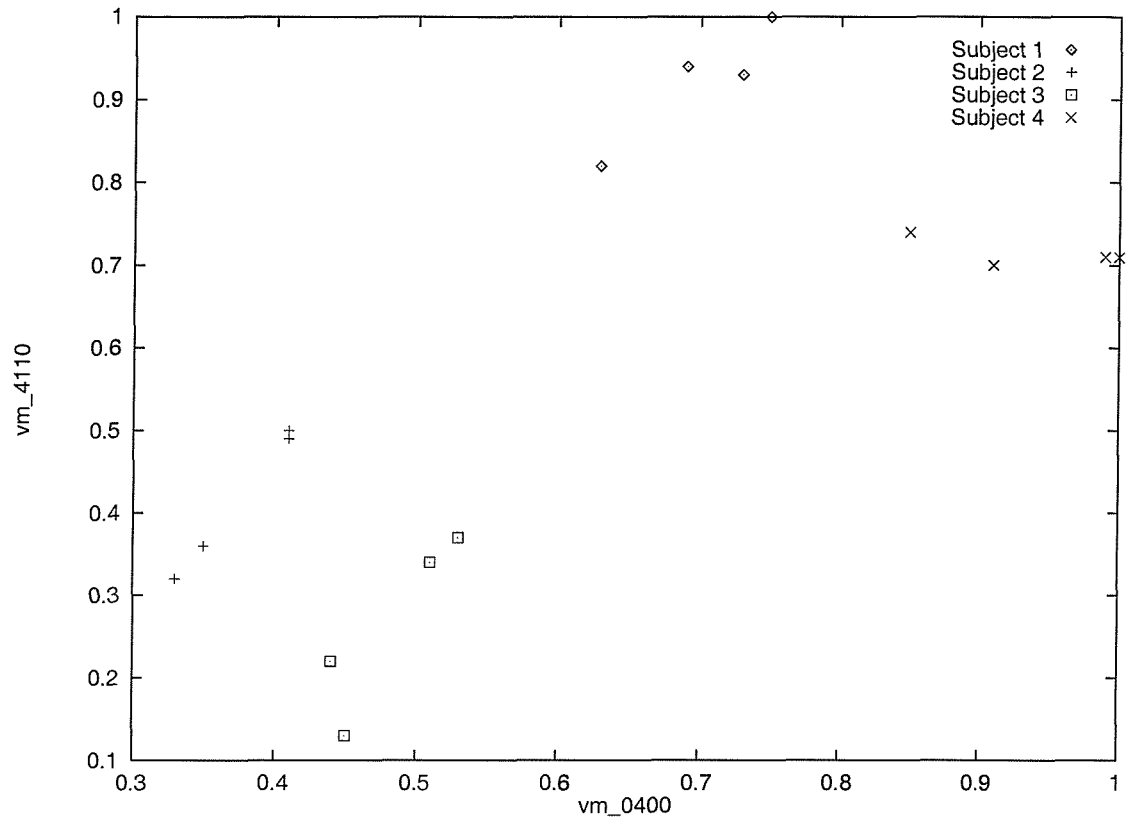


(a) 3 velocity moments - vm_{2310}, vm_{2320} and vm_{0200}.

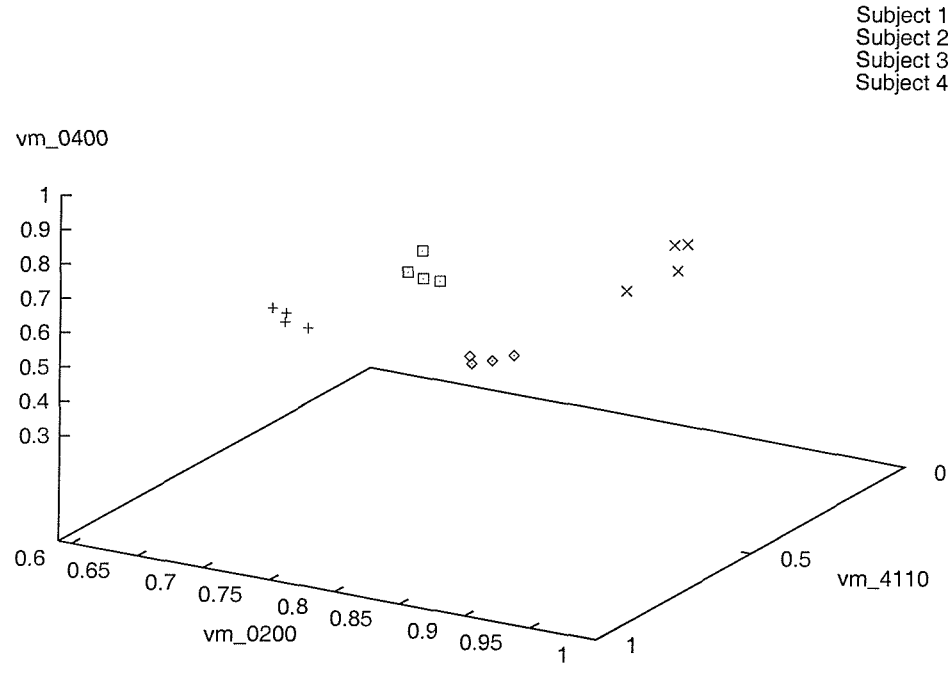


(b) 2 velocity moments - vm_{2310} and vm_{2320}.

Figure 5.3: Normalised ST classification results for the SOTON database.



(a) 2 velocity moments - vm_{0400} and vm_{4110}.



(b) 3 velocity moments - vm_{0200}, vm_{0400} and vm_{4110}.

Figure 5.4: Scatter plot for the TTs from the SOTON database.

5.2.2 *UCSD database*

Here we analyse the UCSD subject database (captured at the University of California San Diego), as used by Little [46] and Huang [29]. It consists of six subjects with seven sequences per subject. In each sequence the subjects are walking (outdoors) from right to left, along a slight incline in front of a static background, an example of which can be seen in Figure 5.5. The subjects were first extracted using the statistical technique, as detailed in Section 4.4.2 and Appendix B. The resulting STs for one complete gait cycle were then windowed using each subject's average velocity, the values of which form part of the database. Using these STs a set of TTs were then generated using the dense optical flow technique described in Section 4.4.4. Example STs and TTs can be seen in Figure 5.6.

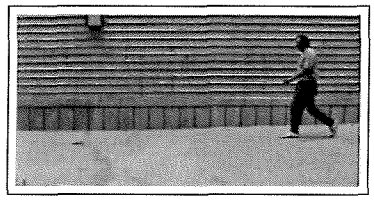


Figure 5.5: Example image from the UCSD database.

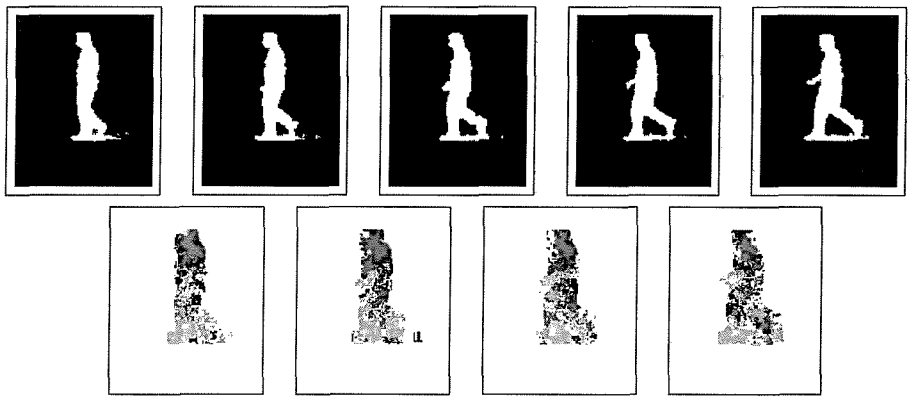


Figure 5.6: Example windowed STs (top) and TTs (bottom) from the UCSD database.

Template	$F > F_{crit}$	$F >> F_{crit}$
ST	74 % (58 %)	4 % (0.8 %)
TT	67 % (51 %)	2 % (0 %)

Table 5.6: Percentages of total Cartesian velocity moments (for the UCSD database) calculated that show promising properties for classification, as determined by the 1 % F_{crit} value of 3.57 - those including velocity information are shown in brackets.

As with the SOTON database 125 velocity moments were calculated on each set of templates, allowing for the windowed data by using the method described in the previous section. Table 5.6 displays the number of moments (in terms of percentages of the total) that showed promising properties with respect to subject clustering

and separation, determined via the F statistic. Even though initial results produced more moments with $F > F_{crit}$ than the SOTON database (comparing Tables 5.1 and 5.6), further examination reveals that less satisfy $F >> F_{crit}$. Table 5.6 shows that only 0.8 % of the STs with $F >> F_{crit}$ ($F > 30$) include velocity. The Scheffe post-hoc analysis allowed the manual selection of two moments (vm_{0310} and vm_{0400}), with the corresponding F_s results shown in Table 5.7. Even though only two moments have been isolated a high classification rate of 80.95 % is achieved, as shown in Table 5.9. The corresponding F statistic values for these two moments can be seen in Table 5.8a. The Scheffe results demonstrate that the two moments complement each other, although particular subject comparisons - (2,5), (2,6) and (5,6) - cause confusion, a result which is reflected in their scatter plot shown in Figure 5.7. (An alternative feature set may produce improved separation for these comparisons).

Analysing the TTs in the same manner produces the classification results shown in Table 5.10, which are plotted in Figure 5.8. Figure 5.8a demonstrates subject clustering, while rotating it about the horizontal plane illustrates the linear relationship between the two of the selected velocity moments (Figure 5.8b). Combining the template descriptions to produce a description of both shape and temporal motion does not improve the classification rate. The classification results for this case can be seen in Table 5.11. Here the two ST velocity moments from Table 5.9 have been combined with one of the TT moments (vm_{0200}).

5.2.3 Discussion - Cartesian velocity moments

The results presented for the SOTON database are encouraging, with high classification rates on a small sixteen sequence database. The inclusion of velocity was seen to alter the subject mean cluster value, as illustrated by the Scheffe post-hoc analysis. This result is dependent on the correlation between the motion and the image sequence, and on the consistency of each subject's multiple image sequences. The STs are describing the subject's overall shape change with respect to their forward motion. In contrast, the TTs are describing limb movement independent of the subject's forward motion (removed by the windowing). By including the average forward velocity into the TT moment calculation, the DC value that is lost by the windowing process can be restored. This produces a range of possible features which can include limb movement, overall body motion, or a combination of the two. If the data was not windowed the TT would contain a combination of overall motion and limb movement, which could not be separated.

The classification results for the UCSD database are somewhat disappointing. Velocity moments with F values greater than the critical values were still produced. However, the amount available for classification was far fewer than for the SOTON database. This may be a combined effect of the reduced resolution of the UCSD

	2	3	4	5	6
1	8.7*	13.0*	0.6	12.3*	22.7*
2		0.4	13.7*	0.3	3.3
3			19.0*	0.0	1.3
4				18.2*	30.5*
5					1.6

(a) vm_{0400}

	2	3	4	5	6
1	6.0*	33.9*	12.9*	4.3*	5.4*
2		11.4*	1.3	0.1	0.0
3			5.0*	14.0*	12.3*
4				2.3	1.6
5					0.1

(b) vm_{0310}

Table 5.7: Scheffe post-hoc results (F_S) for the Cartesian velocity moments - UCSD STs. A * indicates a value greater than the 1 % F_{crit} value of 3.57.

Moment	F-value
vm_{0400}	53.01
vm_{0310}	36.86

(a) STs.

Confidence	F_{crit}
5 %	2.48
1 %	3.57

(b) F_{crit} values.

Moment	F-value
vm_{0200}	81.01
vm_{0400}	52.24
vm_{0310}	9.04

(c) TTs.

Table 5.8: F and F_{crit} values for the selected Cartesian velocity moments on the UCSD database.

Cartesian velocity moments	Classification	
	$k=1$	$k=3$
vm_{0400}	42.86 %	42.86 %
vm_{0400}, vm_{0310}	76.19 %	80.95 %

Table 5.9: Cartesian velocity moments - the UCSD classification results for the STs.

Cartesian velocity moments	Classification	
	$k=1$	$k=3$
vm_{0200}	52.38 %	30.95 %
vm_{0200}, vm_{0400}	64.29 %	52.38 %
vm_{0200}, vm_{0400}, vm_{0310}	54.76 %	57.14 %

Table 5.10: Cartesian velocity moments - the UCSD classification results for the TTs.

Cartesian velocity moments	Classification	
	$k=1$	$k=3$
vm_{0400}(ST), vm_{0310}(ST), vm_{0200}(TT)	76.19 %	76.19 %

Table 5.11: Cartesian velocity moments - the UCSD classification results for the combined templates.

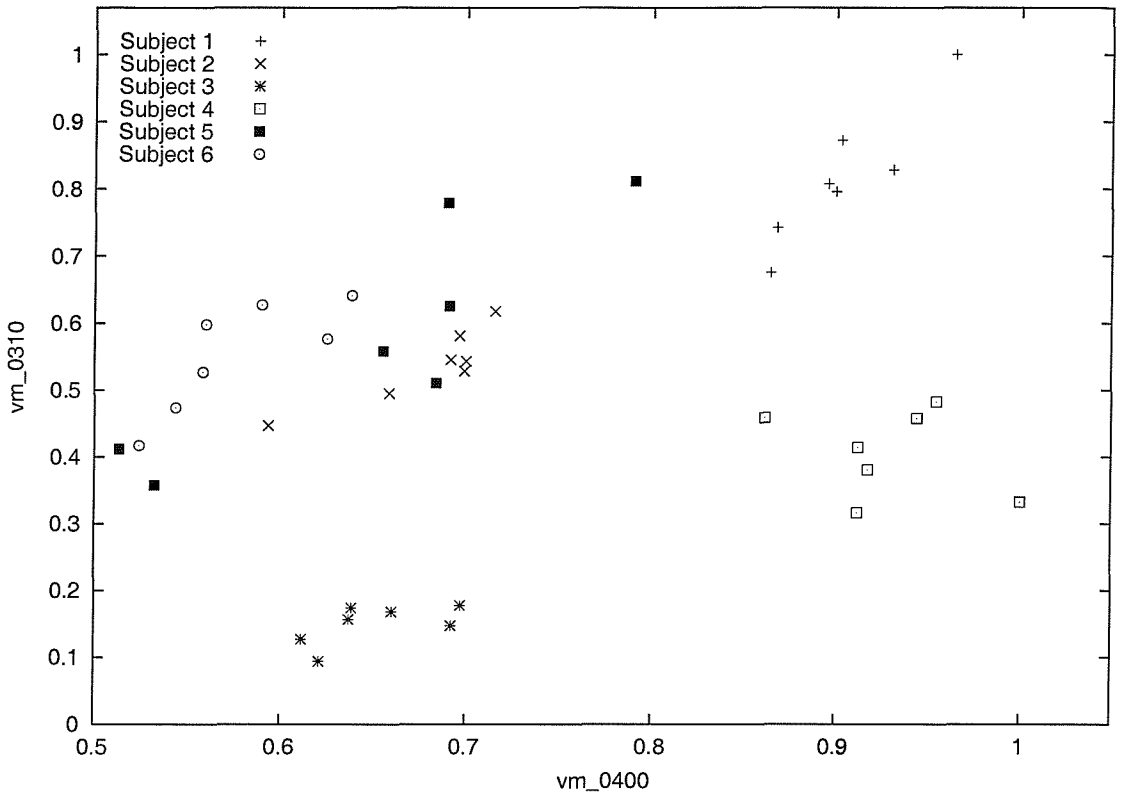
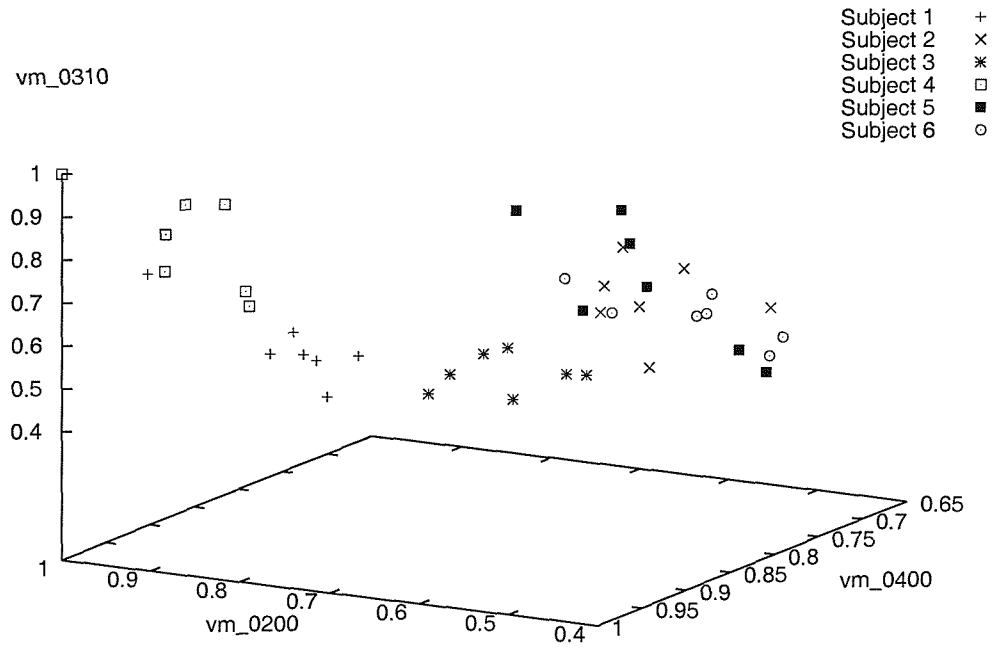
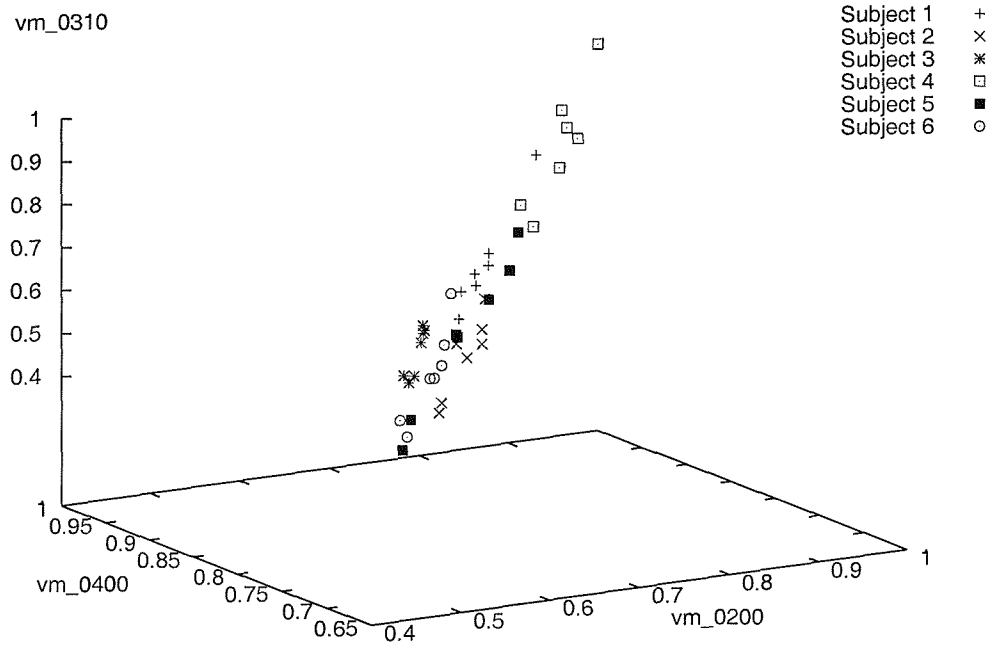


Figure 5.7: Cartesian velocity moments - UCSD ST scatter diagram.

database, its increased size, and the highly correlated descriptors produced by the Cartesian velocity moments. Further to this, there are inconsistencies evident in this database. The subjects in the UCSD database are all walking at similar speeds, but variations will exist within a subject's sequence, implying between-sequence variation (for the same subject). The distance between the camera and the subject varies between some sequences thus the need for scale invariance within the feature description. The current formulation of the Cartesian velocity moments includes sequence-scale invariance (refer to Equation 3.7), however individual image scale invariance is not handled. Here the image-by-image scale invariance version of the Cartesian velocity moments (Section 3.2.2) was not used, so as to avoid further increasing the correlation of the features. The variation in distance leads to interaction between both the ground and the background causing shadows to appear/disappear. This is in addition to the shadows appearing (on most sequences) on the floor between the subject's legs, evident in Figure 5.6. There is also evidence of interaction between some subject's clothes and the background affecting the feature (subject) extraction. These inconsistent characteristics of the database will also affect the Cartesian velocity moment calculations. There may also be secondary effects in the UCSD TTs due to the lower image resolution. The subjects within the UCSD sequences are typically 94 pixels in height, as compared to the



(a)



(b)

Figure 5.8: Scatter plot for the TTs from the UCSD database, showing 3 velocity moments from two different views.

SOTON database of 156 pixels ($\simeq$ 40 % reduction). This reduction in resolution will affect the performance of the TT algorithm; eg. the between-subject motion variations may be of a sufficiently small scale to be lost due to the loss in image resolution. (It is interesting to note that even at the larger scale using the subject's face for classification would appear difficult).

Further to these problems, both databases (SOTON and UCSD) consist of windowed data, which may effectively be reducing the accuracy of the information in which we are interested. Additionally, the subject speed inconsistencies already mentioned may degrade this information. The next section addresses the problem of the highly correlated description by applying the Zernike velocity moments to the gait description, beginning initially with the UCSD database. Further, these moments include individual image scale invariance. It then continues to analyse a larger non-windowed database.

5.3 Zernike velocity moments

The Zernike velocity moments (Equation 3.14) are designed to capture both spatial and temporal information from an image sequence. This is achieved by utilising the orthogonal shape description provided by the Zernike polynomials, defined within the unit disc. Each image within the sequence is first mapped onto the unit disc, and this structural information is then combined with the motion information from between consecutive images. As a direct result of the Zernike polynomials, the individual image descriptions are less correlated and smaller in magnitude. Here we apply these velocity moments to describing gait sequences, with the aim of producing less correlated and more compact descriptions as compared with the Cartesian velocity moments.

5.3.1 Subject mapping

When the subject is mapped onto the unit disc (prior to the Zernike moment calculation for each image), care must be taken to ensure that no part of the subject's shape falls on the perimeter of, or outside the unit disc. The value of β for Equation 2.58 is set, so that the mapped pixels' coordinates are within 90 % of the unit disc's radius. This is done to reduce the effect of the converging polynomials as r approaches unity, illustrated in Figure 2.6. Due to the nature of the encoding of information in the Zernike polynomial (Section 2.4.2), the Zernike moments will efficiently describe the extremities of the subject as they move. Details including the head, arms and legs will appear closer to the perimeter of the unit disc mapping than the torso. This means that the characteristics that are most likely to vary between subjects (i.e. leg, arm and head shape/movement) are described efficiently, whereas details including explicit torso shape will not be as efficiently encoded.

5.3.2 *UCSD database*

The analysis detailed in this section is applied to the UCSD subject database (both STs and TTs), which has already been described in Section 5.2.2. Prior to the translation and scale invariance mapping detailed in Section 2.4.2, the COMs are calculated to adjust the velocity calculations for differences between the average velocity and the actual velocity between successive images. Identical to the approach taken for the UCSD database in Section 5.2.2).

Zernike velocity moments up to order $m, n = 12$, $\mu = 4, \gamma = 0$ (a total of 196) were calculated for all the sequences of STs in the database. The moment orders were chosen in-line with the discussion presented in Section 4.7. To further reduce the size of the selection problem only the magnitudes of the velocity moments were studied. Phase information (of complex Zernike moments) has been shown to be insignificant, in terms of classification, especially when high order moments are included, [37]. Suitable moments for classification were then selected using the one-way ANOVA technique. Moments producing F values below the 1 % confidence level were then removed from the list. Table 5.12 shows the numbers of moments (percentages taken from 196) that showed promising attributes in terms of subject separation and clustering - as determined by the F statistic. Those velocity moments including motion information are shown in brackets. It can be seen that only 5 % of the moments of the STs did not exceed the 1 % threshold. As before $F >> F_{crit}$ indicates $F > 30$. Analysis using the Scheffe post-hoc test revealed that

Template	$F > F_{crit}$	$F >> F_{crit}$
ST	95 % (71 %)	35 % (18 %)
TT	91 % (67 %)	15 % (1 %)

Table 5.12: Percentages of total Zernike velocity moments (for the UCSD database) calculated which show promising properties for classification, as determined by the 1 % F_{crit} value of 3.57 - those including velocity information are shown in brackets.

complete separation of the STs (in terms of classes, or subjects) could be achieved using five of the velocity moments from the remaining list. The corresponding F statistic values for the manually selected moments can be seen in Table 5.14a, along with the Scheffe results in Table 5.13. (The use of parentheses for the moment order i.e $A_{(12)220}$ is to disambiguate between moment orders that are > 9). Here the entries in the Scheffe tables are the F_S values for the subject pairs indicated by the row and column. The combination of the five velocity moments separates all the subject combinations above the 5 % confidence level (all except two combinations are separated above the 1 % confidence level). Prior to classification the velocity moments were normalised by their maximum values, to ensure that moments with larger average values did not bias the results. Table 5.15 details the classification

	2	3	4	5	6
1	32.7*	1.8	63.4*	0.23	3.3
2		19.0*	5.1*	27.4*	15.1*
3			43.6*	0.7	0.2
4				56.0*	37.8*
5					1.9

(a) A_{8210}

	2	3	4	5	6
1	2.6	10.6*	11.7*	2.7	2.1
2		2.8	3.3	0.0	0.0
3			0.0	2.6	3.3
4				3.1	3.9*
5					0.0

(b) $A_{(12)220}$

	2	3	4	5	6
1	4.1*	7.8*	0.9	2.4	0.0
2		0.6	1.2	0.2	5.2*
3			3.5	1.5	9.1*
4				0.4	1.4
5					3.2

(c) $A_{(12)420}$

	2	3	4	5	6
1	27.5*	17.3*	30.5*	2.4	5.9*
2		1.2	0.0	13.5*	8.0*
3			1.8	6.8*	3.0
4				15.7*	9.6*
5					0.7

(d) A_{5100}

	2	3	4	5	6
1	4.9*	43.3*	18.5*	3.0	0.6
2		77.3*	42.4*	15.5*	9.0*
3			5.2*	23.6*	33.5*
4				6.6*	12.3*
5					0.9

(e) A_{9900}

Table 5.13: Scheffe post-hoc results (F_S) for the Zernike velocity moments - UCSD STs. A * indicates a value greater than the 1 % F_{crit} value of 3.57.

Moment	**F-value**
A_{8210}	102.73
$A_{(12)220}$	16.25
$A_{(12)420}$	13.87
A_{5100}	48.02
A_{9900}	98.81

(a) STs.

Confidence	F_{crit}
5 %	2.48
1 %	3.57

(b) F_{crit} values.

Moment	**F-value**
$A_{(10)200}$	152.11
A_{9100}	55.74

(c) TTs.

Table 5.14: F and F_{crit} values for the selected Zernike velocity moments on the UCSD database.

Zernike velocity moments	Classification	
	$k=1$	$k=3$
A_{8210}	61.90 %	52.38 %
A_{8210}, $A_{(12)220}$	80.95 %	76.19 %
A_{8210}, $A_{(12)220}$, $A_{(12)420}$	85.71 %	88.10 %
A_{8210}, $A_{(12)220}$, $A_{(12)420}$, A_{5100}	97.62 %	97.62 %
A_{8210}, $A_{(12)220}$, $A_{(12)420}$, A_{5100}, A_{9900}	100.00 %	100.00 %

Table 5.15: The Zernike velocity moments - UCSD classification results for the STs.

Zernike velocity moments	Classification	
	$k=1$	$k=3$
$A_{(10)200}$	54.76 %	35.71 %
$A_{(10)200}$, A_{9100}	97.62 %	95.24 %

Table 5.16: The Zernike velocity moments - UCSD classification results for the TTs.

Zernike velocity moments	Classification	
	$k=1$	$k=3$
$A_{(10)200}$ (TT), A_{9100} (TT), A_{8210}(ST)	100.00 %	100.00 %

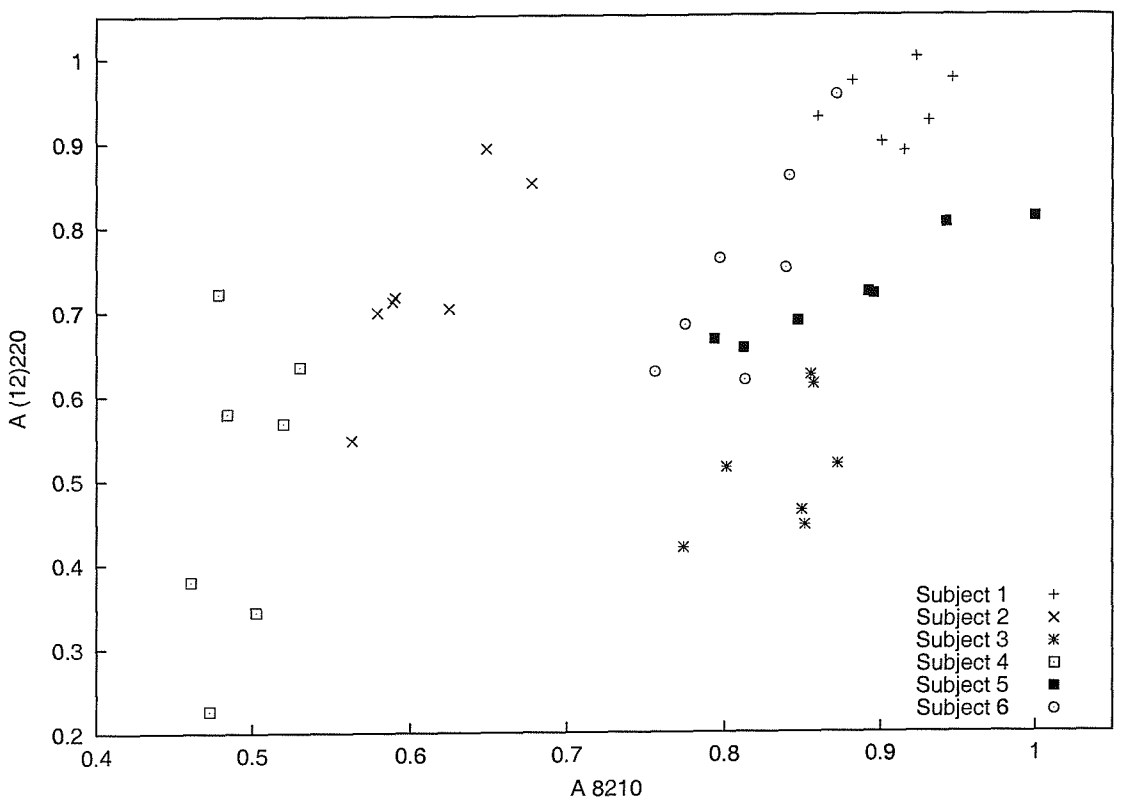
Table 5.17: Combining templates - UCSD classification results for the Zernike velocity moments.

results. It can be seen that a classification rate of over 80 % with $k = 1$ is achieved using only two features. 100 % classification is achieved using just five features, for both $k = 1$ and $k = 3$. Figure 5.9a shows a scatter plot of the first two Zernike velocity moments, while Figure 5.9b shows a scatter plot of the first three moments illustrating both clustering and cluster separation.

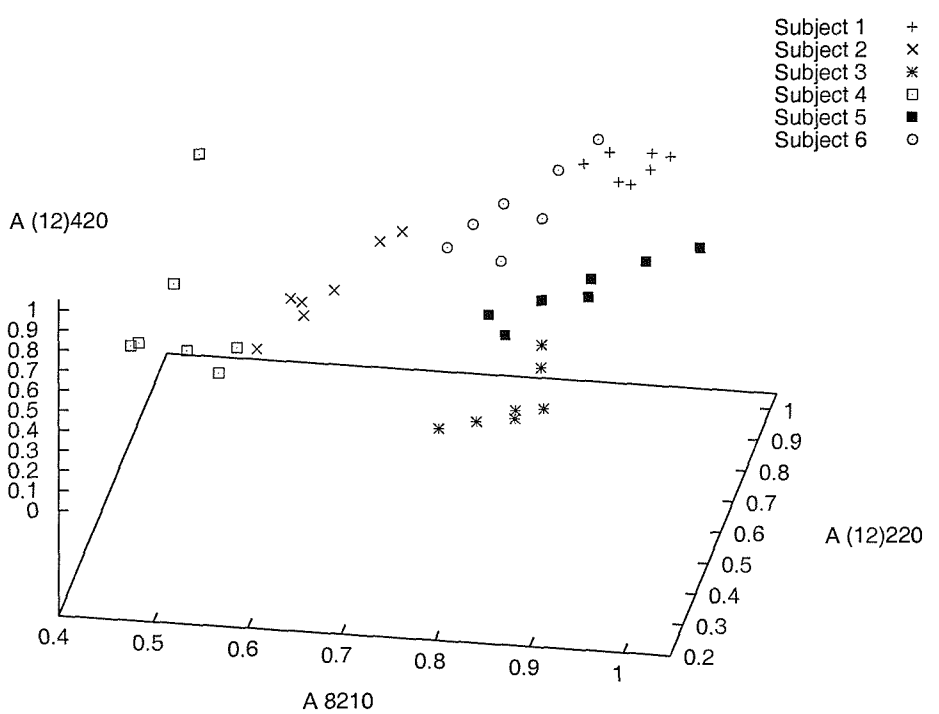
Applying the same analysis to the TTs of the UCSD database produced the classification results shown in Table 5.16, resulting in a classification rate of 95 % using two velocity moments, displayed in Figure 5.10. Here velocity moments which contain predominately structural information appear useful. The corresponding F statistic values can be seen in Table 5.14c. By then adding a single ST velocity moment (A_{8210}) to the TT moments, 100 % classification is achieved (shown in Table 5.17), improving previous results on the same database, Section 5.2.2 and [46], and with fewer descriptors. By combining the two types of templates, a description of both shape with motion (STs) and an individual's limb motion (TTs), (independent of their average motion) is produced. Naturally, a larger database would doubtless require more moments to separate subjects, but it is perhaps worth noting that only a basic classifier has been used.

5.3.3 CMU databases

There are four CMU databases (captured at the Carnegie Mellon University), each consisting of STs of the same 25 subjects walking on a treadmill. The computation



(a)



(b)

Figure 5.9: Scatter plots of the selected Zernike velocity moments used for classification of the UCSD STs.

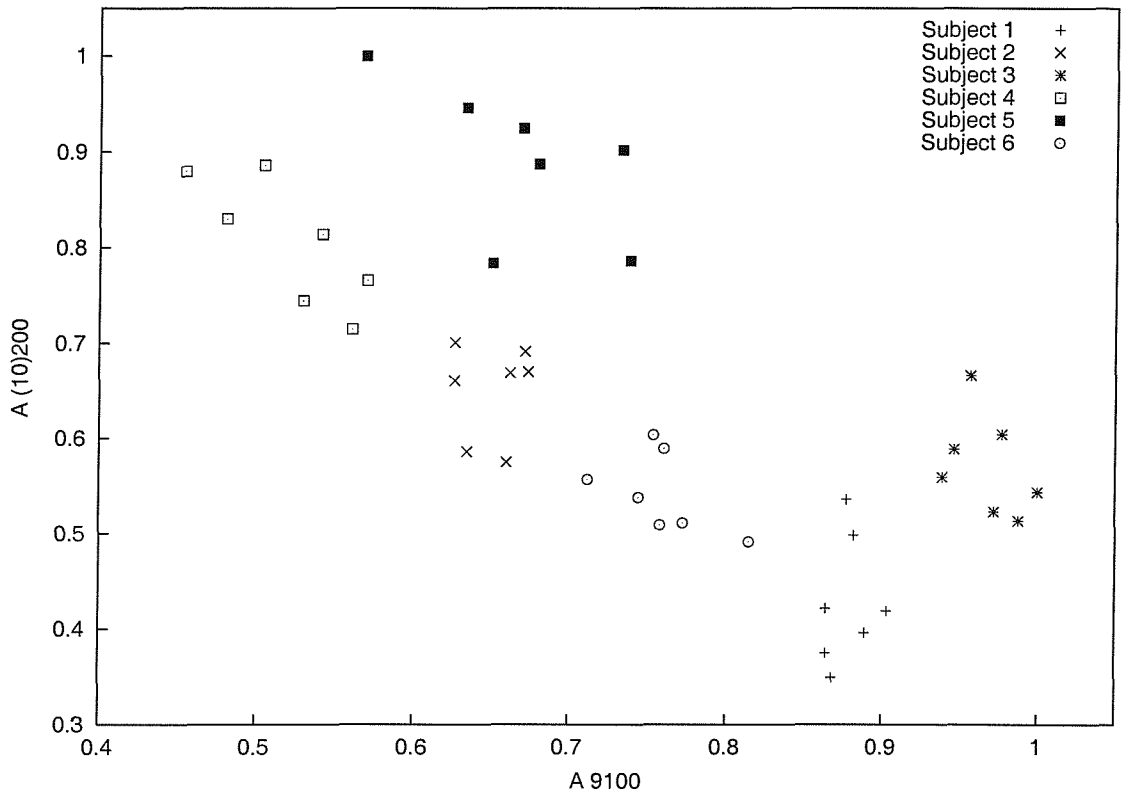


Figure 5.10: Scatter plot of the selected Zernike velocity moments used for classification of the UCSD TTs.

CMU Database	Camera	Walking speed	Viewing angle
CMU_03_7_s	03_7	slow	normal
CMU_03_7_f	03_7	fast	normal
CMU_05_7_s	05_7	slow	oblique
CMU_05_7_f	05_7	fast	oblique

Table 5.18: The different CMU databases.

of the TTs was not possible due to most of the original data being unavailable. Two of the databases view the subject from the side (normal to the camera), the remaining two view from an oblique angle ($\simeq 45^o$ from normal), as shown in Figure 5.11. For each viewing angle there is data captured at two different walking speeds, classified as slow and fast. Table 5.18 summarises these details. The STs were generated from the original colour data using a simple background subtraction technique. The result was median filtered (3×3 mask) to remove the effects of noise caused by variations in lighting, etc. Figure 5.12 shows an example image from one of the CMU databases, along with its corresponding ST. All subjects have four sequences per database of them walking for one complete gait cycle, heel strike to heel strike of the same foot, producing STs with no forward or DC velocity. However, fluctuations about the mean x position may exist.

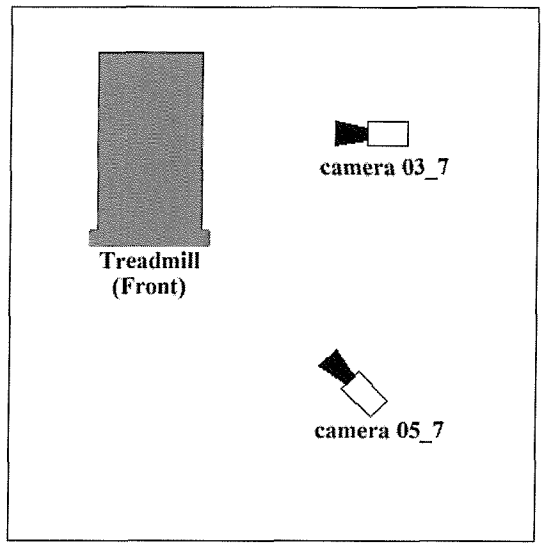
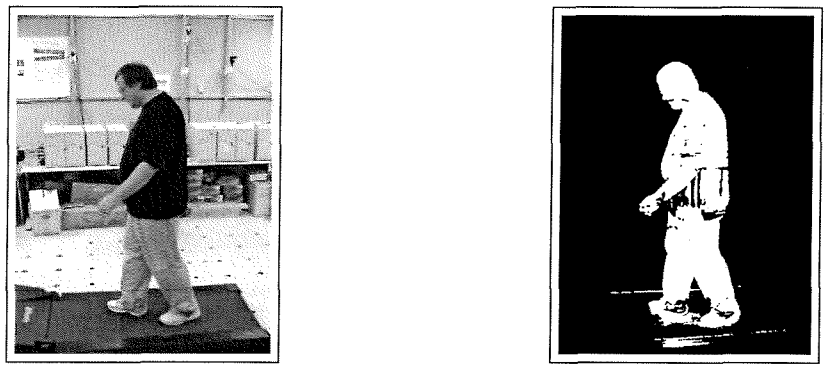


Figure 5.11: A plan view of the treadmill and cameras for the CMU databases.



(a) (b)

Figure 5.12: Example image from the CMU_03_7_s database (a) and the corresponding ST (b).

For the first database of STs ($CMU\_03\_7\_s$) a total of 784 Zernike velocity moments were calculated (on 100 sequences in total). This took approximately ten days to process on a cluster of eight 1 GHz machines. Consequently, this list was reduced to those moments which satisfied $F > 30$, along with those moments which proved useful in the UCSD analysis. This reduced the moment list for the remaining three databases to 90, 43 of which included velocity information (both x and y). Table 5.20 shows the k-nn classification results for each database using all 90 velocity moments, showing results of over 90 % for all four databases. A further reduced feature set was achieved using the one way ANOVA technique with the Scheffe post-hoc tests. The F statistic results for the manually selected moments can be seen in Tables 5.19a and b, all of which are much greater than the F_{crit} values shown in Table 5.19c. The moments used (for both camera views) to classify the fast and slow walks are identical, while between camera views they differ. Table 5.21 shows the classification results for this manually refined list of 6 velocity moments,

all of which are over 85 %.

Due to the nature of the treadmill, none of the ST sequences have any forward velocity information. This is apparent in the selected velocity moments (refer to Table 5.21), as none include a forward velocity term i.e. $A_{**?*}$. If the treadmill speed for each subject had been known, then this could have been used as a DC value in the velocity moment calculation. This would be supplemented by any x direction variations present in the STs (similarly achieved for the SOTON and UCSD databases, refer to Section 5.2). However, a subject's y direction motion information is visible in treadmill data, appearing as a vertical 'bobbing' motion as they walk. This richness of y direction motion information is reflected in the selected moments, as many of them include y velocity information (mostly magnitude information i.e. A_{***2}). In both cases the fast walk sequences have lower classification results as compared with the slow walk sequences, reflecting a loss in temporal resolution i.e. less images describing the gait cycle. It is interesting to note that these STs are richer in information than their UCSD versions. Not only due to their increased resolution, but also due to the large number of holes within each subject's perimeter, as shown in Figure 5.12b. The holes correspond to areas of the background that have interfered with the background subtraction. This has effectively increased the amount of spatial information as these holes are correlated to the subject's shape and movement. One study exploited such a technique to increase the efficiency of moment descriptors, by occluding part of the shape being described [76]. In this work a shape is occluded using a set of circles, producing a family of shapes which represent the original object. This results in (potentially) more spatial information being available, reducing the need, in terms of classification, for higher order moments as the shape essentially becomes more unique.

5.3.4 HiD database

The HiD database (Human ID at a distance research program) used here consists of 50 subjects, with 4 sequences of each subject, a total of 200 sequences ($\simeq$ 6000 images). The subjects are walking around a continuous bone-shaped track, the main shank of which is normal to the camera. The sequences studied here contain the subjects walking from left to right for one and a half gait cycles (three consecutive heel strikes). The subjects are walking in a relaxed manner, achieved by letting them settle into their walk (around the continuous track) before filming. The surface of the track was flat, while the loops at either end of the track are out view of the camera, allowing the subject to be walking in a straight trajectory when normal to the camera. Figure 4.4 shows a diagram of the main shank of the track minus the end loops. Due to the background and the controlled lighting conditions, chroma-key

Moment	F values	
	Slow	Fast
A_{8202}	98.49	63.50
$A_{(12)400}$	114.47	133.94
A_{2000}	151.30	200.86
A_{2200}	155.99	138.75
$A_{(11)(11)01}$	40.47	34.64
A_{4002}	87.63	82.97

(a) CMU_03_7 F values.

Moment	F values	
	Slow	Fast
A_{8200}	271.85	28.69
$A_{(12)402}$	141.09	93.53
A_{2200}	247.34	81.95
A_{4202}	249.37	164.94
A_{7700}	116.43	64.44
A_{4402}	296.24	223.53

(b) CMU_05_7 F values.

Confidence	F_{crit}
5 %	1.66
1 %	2.05

(c) F_{crit} values.

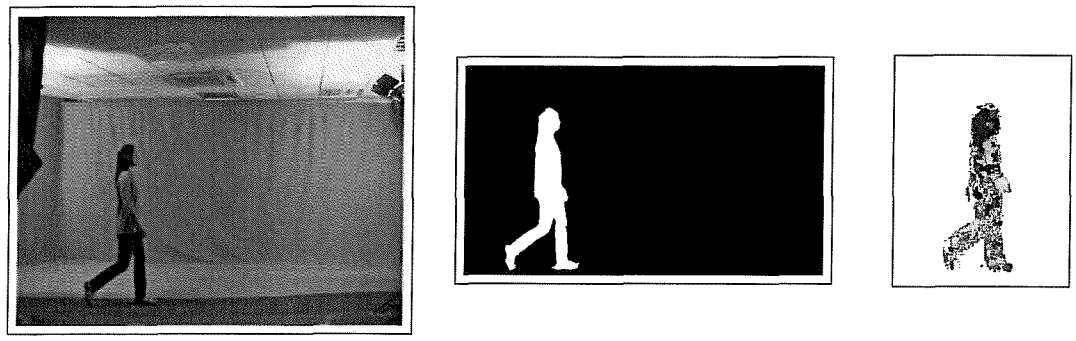
Table 5.19: F and F_{crit} values for the selected Zernike velocity moments on the CMU databases.

Camera	Classification $k = 1$		Classification $k = 3$	
	Slow	Fast	Slow	Fast
CMU_03_7	100.00 %	100.00 %	100.00 %	100.00 %
CMU_05_7	100.00 %	99.00 %	99.00 %	95.00 %

Table 5.20: The classification results for the four CMU databases using 90 velocity moments.

Camera	Zernike velocity moments	Classification $k = 1$		Classification $k = 3$	
		Slow	Fast	Slow	Fast
CMU_03_7	A_{8202},$A_{(12)400}$,A_{2000} A_{2200},$A_{(11)(11)01}$,A_{4002}	91.00 %	91.00 %	90.00 %	87.00 %
CMU_05_7	A_{8000},$A_{(12)402}$,A_{2200} A_{4202},A_{7700},A_{4402}	95.00 %	96.00 %	92.00 %	87.00 %

Table 5.21: The classification results for the four CMU databases using 6 velocity moments.



(a) Original (b) ST (c) TT

Figure 5.13: Example image from the HiD database, its corresponding cropped ST and TT (computed from image n, $n+1$).

extraction was possible, detailed in Section 4.4.3. Both STs and TTs were computed for this database. Figure 5.13 shows an example image from the database along with its corresponding cropped ST and TT (histogram equalised and computed from image n, $n+1$). Due to the increased resolution of the images and the distance over which the subjects walked, the optical flow for the TTs was computed within a moving window, moving at the subject's average velocity, a method already used in gait recognition by Little [45]. The average velocity was calculated using the COM information from the STs. As before the TTs are (effectively) windowed data, so the average velocity is placed back into the velocity moment calculation (as done for the SOTON and UCSD databases). However, due to the increased resolution and size of the HiD TTs dataset, the Zernike moment scaling (for the TTs) was switched off to avoid problems through scaling large, non-binary images (i.e. the mapping could scale the subject causing it to exceed the unit disc's area). The images were instead scaled to appear visually central to the unit disc i.e. the thresholded COM was used in the mapping (in-place of the actual greyscale COM).

A list of 234 Zernike velocity moments was computed on the STs and TTs. The moment list contained moments describing both x and y velocity components along with spatial information. The list was manually constructed using the results from previous database analyses and contained moments up to, and including order/rotations $m, n = 12$. This prior selection was made to help reduce the computation time. The k-nn classification results for the complete moment list of 234 moments on the STs and TTs can be seen in Table 5.23. These classification results are low, suggesting the need for feature selection. Results for a subset of eight ST moments selected using the ANOVA technique, are shown in Table 5.24. Table 5.22a summarises the F statistic values for the eight selected ST moments, whereas Table 5.22c shows those for a set of five selected TT moments. The F statistic values shown in Table 5.22a and c are all greater than the critical values shown in Table 5.22b. A high classification of 83.50 % ($k = 3$) is achieved on this large

database using eight ST velocity moments, as shown in Table 5.24. Table 5.25 shows the classification rates for the five selected TT velocity moments which are relatively low in comparison. The selected TT velocity moments favour those holding solely spatial information (i.e. A_{**00}). A similar result was found upon analysis of the UCSD TT database, supporting the hypothesis that the TTs hold detailed information about a subject's limb motion (which may not vary enough between-subjects on its own to allow good subject separation), while the STs hold global shape/motion differences. It is interesting to note that these results (STs and TTs) consistently show the $k = 1$ classification results to be greater than $k = 3$. This suggests that the feature space is closely packed (with respect to subject clusters). There are two obvious solutions to overcome this problem. The first is to increase the dimensionality, using more features to increase cluster separation, or secondly to use a more sophisticated classifier, as mentioned in Section 4.6. Although, this effect may be caused by the normalisation of the moment values. The normalisation is used to stop biasing of the k-nn classifier by moments which naturally produce larger values. However, if one subject produces significantly different feature values to the rest of the database, the remaining subjects within the database (and the differences between them) will be compressed into a small area of the feature space. This is illustrated in Figure 5.14 where 20 subjects from the HiD are plotted, subject 20's features are significantly different from the remaining subjects which are closely grouped. Alternatively the same effect will be observed if an outlier to a subject cluster exists i.e. one of subject 18's sequences in Figure 5.14. (A situation which is explored further in the Section 5.3.5). These results highlight the possible need for an alternative classifier when analysing larger datasets, or in situations where the feature's order of magnitude may vary. Alternatively, we can combine the two template results (as done for the UCSD database). Table 5.26 displays the results of combining the eight selected STs velocity moments with the first four TT velocity moments. This results in a higher classification rate of 96 % ($k = 3$). The proximity of this result to the $k = 1$ classification rate of 97 % suggests that the feature space is less packed with respect to between-subject differences, than it was when using just STs or TTs alone.

5.3.5 Case studies

This section details two simple case studies which help to illustrate that the Zernike velocity moments contain both structural information and motion information, as reflected in their formulation (Equation 3.14). The first of these examples was produced by adding one extra subject to the HiD database, using the same laboratory conditions as used to capture the original database (refer to Sections 4.4.3 and 5.3.4). The subject's data consisted of three sequences of him walking in a normal

Moment	F-value
A_{6000}	61.51
A_{8200}	61.54
A_{8810}	25.97
$A_{(12)(12)20}$	21.26
A_{7110}	23.31
A_{2200}	38.41
A_{8410}	15.15
$A_{(12)400}$	45.77

(a) STs.

Confidence	F_{crit}
5 %	1.44
1 %	1.67

(b) F_{crit} values.

Moment	F-value
A_{5100}	59.67
A_{6200}	71.95
A_{9900}	85.75
$A_{(10)(10)00}$	106.75
A_{6610}	47.03

(c) TTs.

Table 5.22: F and F_{crit} values for the selected Zernike velocity moments on the HiD database.

HiD Template	Classification	
	$k = 1$	$k = 3$
ST	74.00 %	57.50 %
TT	52.50 %	28.50 %

Table 5.23: The classification results for the HiD database using 234 velocity moments.

Zernike velocity moments	Classification	
	$k = 1$	$k = 3$
A_{6000}, A_{8200}	39.50 %	31.00 %
A_{6000}, A_{8200}, A_{8810}	63.00 %	47.50 %
A_{6000}, A_{8200}, A_{8810}, $A_{(12)(12)20}$, A_{7110}	80.00 %	66.00 %
A_{6000}, A_{8200}, A_{8810}, $A_{(12)(12)20}$, A_{7110}, A_{2200}, A_{8410}, $A_{(12)400}$	93.50 %	83.50 %

Table 5.24: The HiD classification results for the spatial templates.

Zernike velocity moments	Classification	
	$k = 1$	$k = 3$
A_{5100}, A_{6200}	24.50 %	17.00 %
A_{5100}, A_{6200}, A_{9900}, $A_{(10)(10)00}$	52.50 %	38.00 %
A_{5100}, A_{6200}, A_{9900}, $A_{(10)(10)00}$, A_{6610}	61.50 %	50.50 %

Table 5.25: The HiD classification results for the temporal templates.

Zernike velocity moments	Classification	
	$k = 1$	$k = 3$
(STs) A_{6000},A_{8200},A_{8810},$A_{(12)(12)20}$,A_{7110},A_{2200},A_{8410},$A_{(12)400}$ (TTs) A_{5100},A_{6200},A_{9900},$A_{(10)(10)00}$	97.00 %	96.00 %

Table 5.26: The HiD classification results for combining the templates.

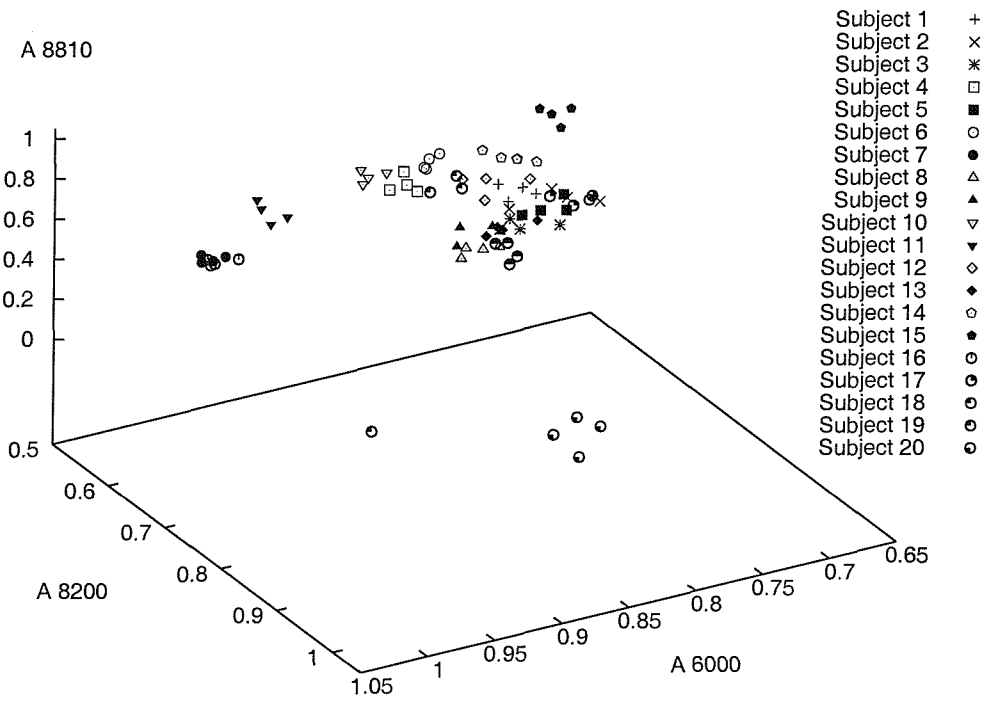


Figure 5.14: 20 subjects from the HiD database plotted for 3 Zernike velocity moments (used for classification), illustrating the possible effects of normalising the features.

relaxed manner, achieved by asking him to walk normally around the continuous track. For his fourth sequence he was asked to walk in an abnormal manner. This fourth sequence produced the subject walking with more vertical motion throughout the sequence, along with variations in stride length and arm motion. (The subject swung his arms considerably more than usual and walked in a 'jerky' manner.) A set of Zernike velocity moments for all four sequences was then calculated, allowing the subject to be added to the database. Figures 5.15 and 5.16 show the results of adding this new subject to the database. To allow for visualisation this extra subject is compared with 9 others (picked at random) from the HiD database. Note these plots show alternative Zernike velocity moments to those already presented for the classification of the HiD database, illustrating the availability of features that produce tight class (or subject) clustering. The scatter plot has been rotated about the horizontal plane in a clock-wise direction to produce the four plots. The additional subject (Subject 10) is represented by the cluster of triangles at the centre of the first plot (Figure 5.15). As the plot rotates the feature points cluster and then separate, resulting in the final plot (Figure 5.16) where the feature point corresponding to the abnormal walk is an outlier, visible at the far left of the plot. It is worth noting that all three of the velocity moments in the plots include motion

information. The second example in this data illustrates a similar result. This subject chose to hold their chin throughout one of their sequences, producing one sequence with little or no arm-swing. The difference caused in the spatial templates is reflected in the feature point clustering, as can be seen in Figures 5.15 and 5.16, where the outlier to Subject 3 is the sequence in question. These velocity moments fail to allow clustering of the four feature points for Subject 3, however, an alternative moment set exhibit improved clustering as shown in Figure 5.17. The x axis (A_{8200}) is a measure of the subject's spatial area and clusters well, which agrees with the subject's arm being visible while holding their chin. A full study of potential within-class variation caused by these factors and others (i.e. same subject with different footwear, carrying objects etc.) is within Southampton's part of the HiD research program and beyond the scope of this thesis. However, within-class variations will also exist due to alternative application environments motivating the next chapter on performance analyses.

5.4 Discussion

5.4.1 Limiting factors

One large constraint on this work is the definition of the gait cycle. The work detailed here has focussed on describing one complete gait cycle. Due to the periodic nature of gait, theory suggests (as applied to the velocity moments) that it will not matter where in the gait cycle the description (or sequence) starts, just as long as one complete cycle is described - the results should not differ. One possible way of automatically determining a complete cycle, would be to study the low order moments of the ST sequence. Here the periodicity of the low order moments should correspond to the periodicity of the gait cycle, thus enabling the length of the sequence to be determined. Figure 5.18 shows the periodic nature of a ST's mass (μ_{00}) as it varies through a gait cycle. The local maxima correspond to the legs at full stride, whereas the minima signify the legs being together (refer to Figure 4.1).

One area which has not been addressed is the problem of image calibration. None of the images used within the databases have been corrected for radial (lens) distortions or colour calibration. (The requisite calibration information was unavailable.) The lens distortions as the person walks across the field of view can be visibly apparent through the image sequences, especially where large resolution images are used, i.e. the HiD database. (This may also be the reason why there is a slight degrading trend in Figure 5.18.) This creates slightly 'warped' silhouettes, which vary depending on where in the field of view the subject is positioned. However, most of the subjects within each database are walking along approximately the same part of each track, thus between-subject distortions may be consistent. This however is not guaranteed. Also, sequences of subjects walking in different

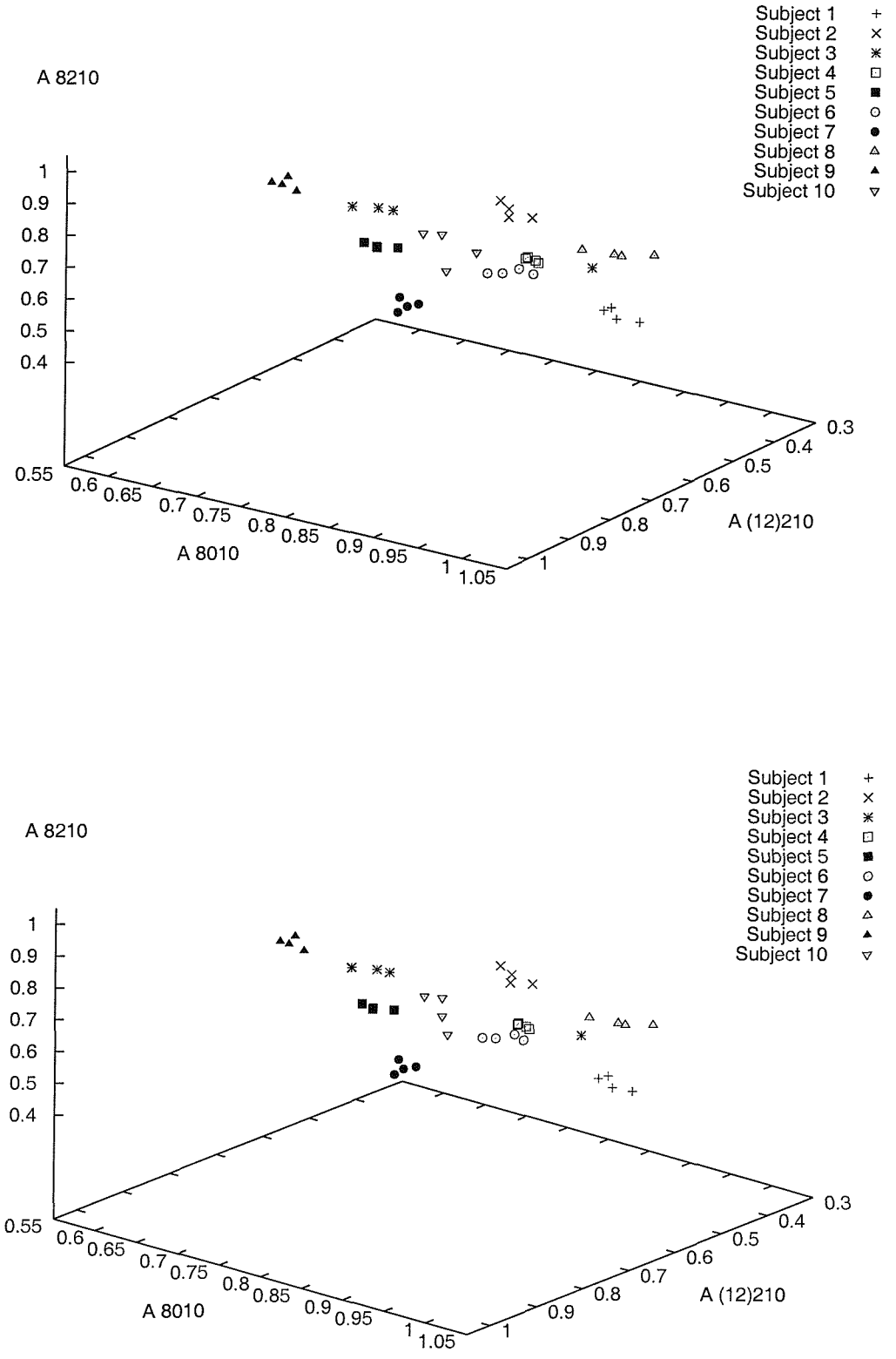


Figure 5.15: The result of an abnormal walk causing one of subject 10's feature points to drift. The 3D scatter plots are of the same three velocity moments, rotated about the horizontal axis, demonstrating the feature point clustering.

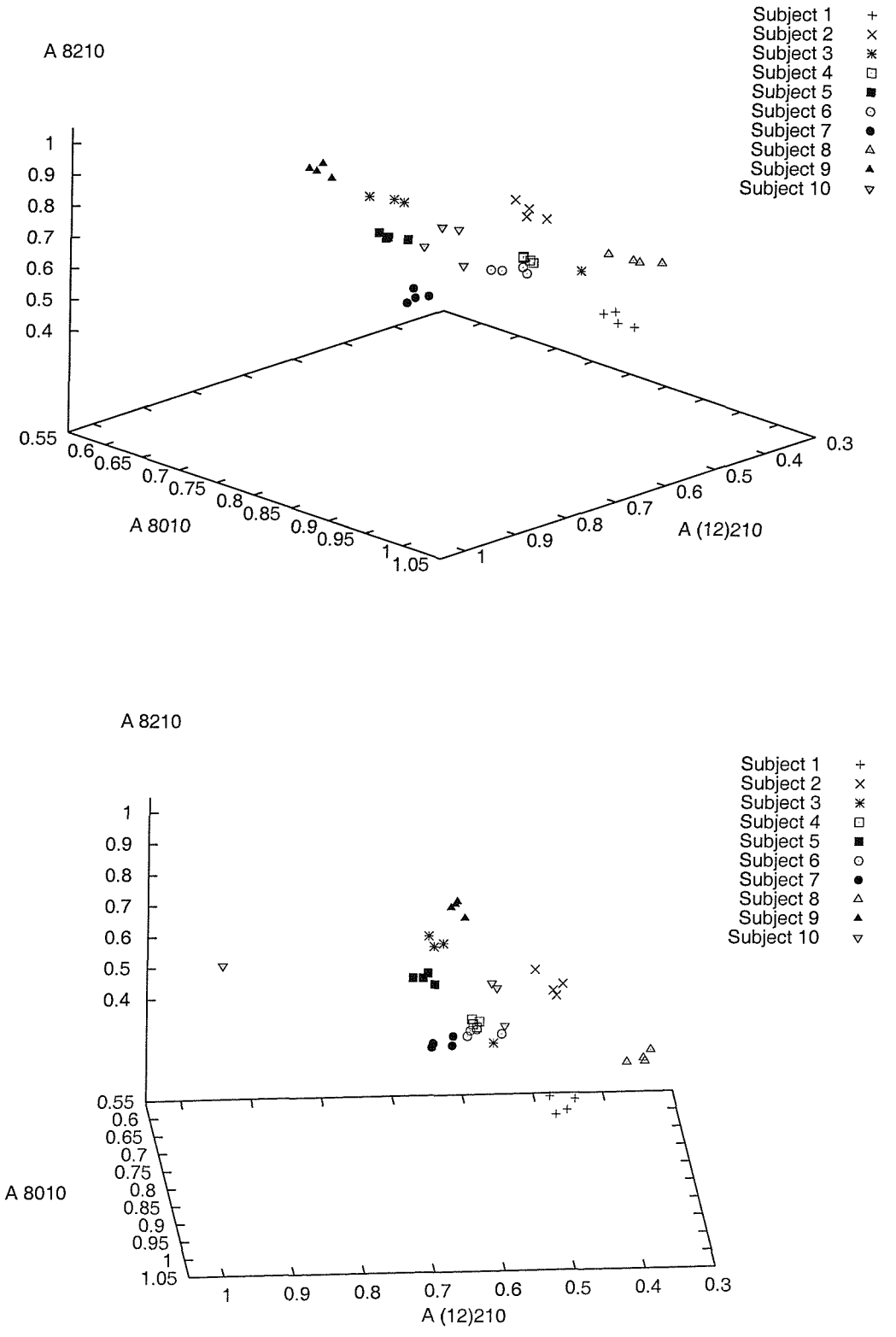


Figure 5.16: The result of an abnormal walk causing one of subject 10's feature points to drift. The 3D scatter plots are of the same three velocity moments, rotated about the horizontal axis, demonstrating the feature point clustering (top) and separating (bottom).

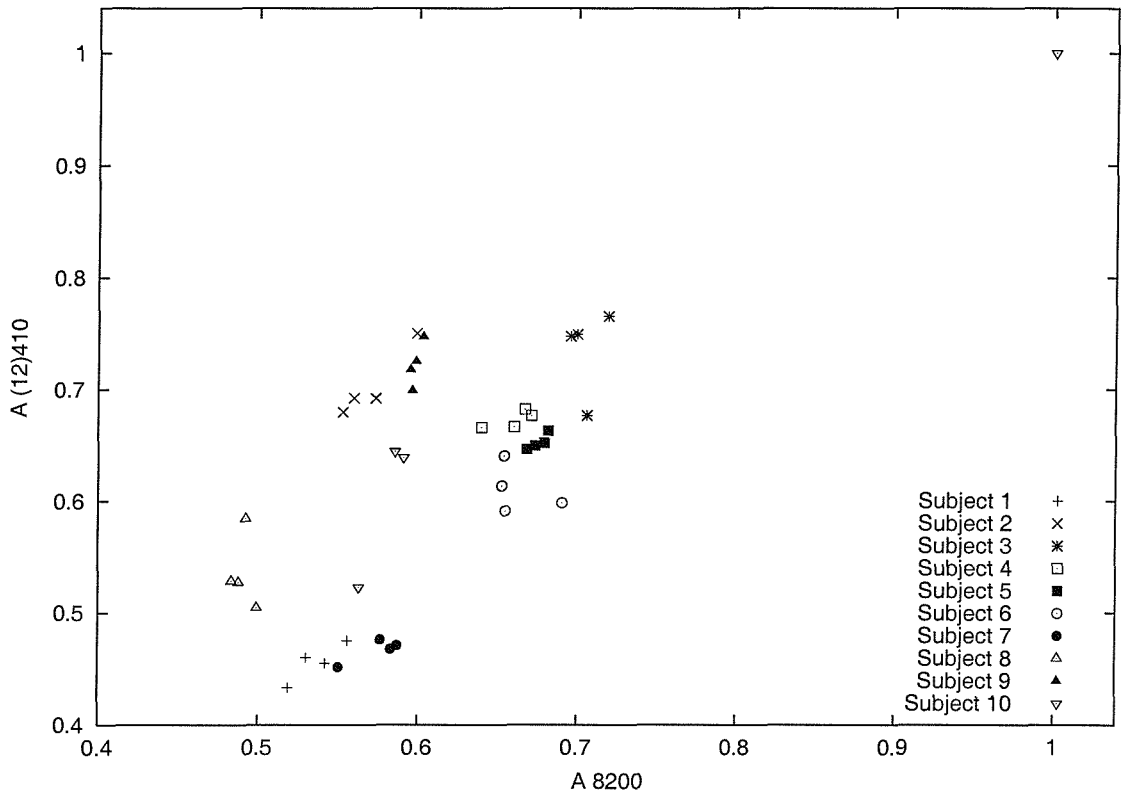


Figure 5.17: Improved clustering for subject 3 who chose to hold their chin through one of their walking sequences.

directions (i.e. the SOTON database) are more likely to experience problems, due to the distortions affecting different halves of the subject's silhouette, depending on their direction of motion. Eg. walking in one direction may cause the leading leg to be distorted, while walking in the opposite direction (along the same part of the track) would cause the trailing leg to be distorted. Thus, a detailed study into direction independent gait classification would require lens corrected data. The problem of colour calibration (in terms of this research) has less of an effect. Variations may be detectable in the optical flow calculations and the subject extraction methods as these depend directly on luminance values. The effects of these calibration issues will be present in the subject (or class) variance of the corresponding features. Correcting these issues, would potentially improve the clustering of the subjects, thus the overall classification results.

Possible improvements could be made to the TT generation. One possibility is to produce higher-detail versions, with increased numbers of displacements removing the need for the average-velocity windowing. Alternatively, noise within the resultant TTs could be reduced by use of connected components analysis to help remove small areas of variation, which may be due to noise (as employed by Little [45]). However, this may in turn remove possible detailed information that

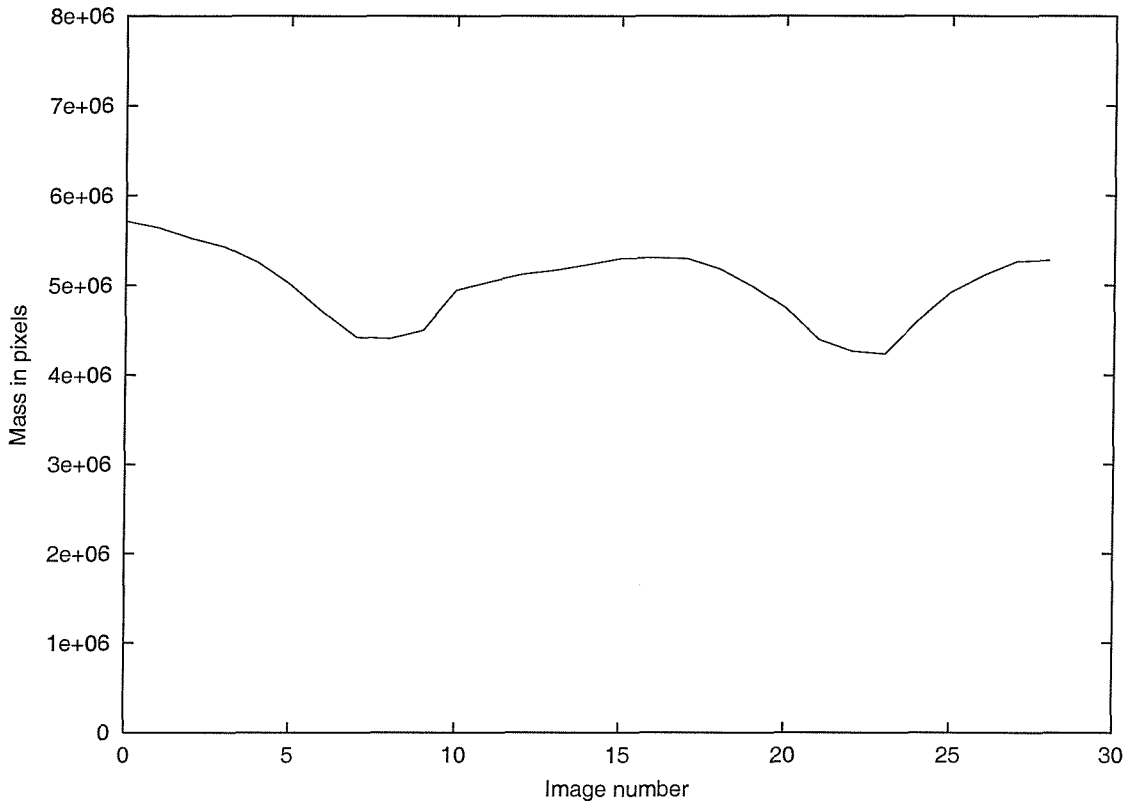


Figure 5.18: Periodic nature of the varying mass (μ_{00}) of a ST sequence (from the HiD database).

exists between-subjects in a large database. One further limiting factor of this investigation is the type of classifier used. The HiD database results (comparing the $k = 3$ and $k = 1$ results) suggest improvements on the classification results could be achieved by using a more complicated classifier. This may more intelligent separation of the subject clusters, improving the performance of the existing selected features. Here, we are primarily interested in the features produced, rather than fine tuning the classification results - a possible route of further investigation.

5.4.2 Symmetry and motion

Previous psychological studies i.e.[13, 18] have concluded that a person's gait can be described in terms of symmetry and that it plays an important part in a persons' gait pattern, producing a synchronous symmetrical pattern of movement [13]. The arms and legs (or pendula) reflect about the y (vertical) axis or torso and are asymmetric about the x (horizontal) axis, further symmetries exist within the leg motion itself [18] (and are inherent to pendulum models) producing symmetry of motion through the temporal sequence. Explicit symmetry operators have already been applied to gait classification [25, 26] producing comparable recognition results (greater than 90 % on a 28 subject database), however, the number of features

utilised is far greater than the results presented here. Alternative area measures may also exploit this symmetry [19, 20]. Symmetry is also an area of interest to the face recognition community eg. [47].

Statistical moments have been shown to produce descriptions that extract symmetric (or asymmetric) characteristics of an image as described in Section 2.2.4. It is apparent that the centralised moments contain information about symmetry. These form the basis of the velocity moments, so it would appear feasible for them to also exhibit these symmetry properties. A direct comparison to Li's results presented in Section 2.2.4 is not possible as they were attained using scale-normalised centralised moments (Equation 2.20), which is not true in the case of the Cartesian velocity moments. This scale normalisation was not used to avoid increasing the (already high) feature correlation. However, some interesting observations can still be made if we consider the Cartesian velocity moments used for classification on the SOTON database. The results presented in Table 5.4 (calculated on the SOTON STs) favour vm_{23**}. On its own the $p = 2$ value (i.e. vm_{2000}) describes the range of pixel spread in the x axis. This can be attributed to the subjects' stride length and range of arm swing, as vm_{2000} is the time averaged version of μ_{20}. Also, on its own the $q = 3$ value (i.e. vm_{0300}) is describing the time averaged skewness in the y axis (or the asymmetry of the distribution in that axis). (It is noted that interaction between these two descriptions will occur for vm_{23**}, producing a correlated time averaged spread with respect to skewness - a fifth order moment.) Further, vm_{2310} will produce a correlated time averaged spread with respect to skewness and motion. All of the moments in Table 5.4 are time averaged (and/or motion weighted) versions, of the low order moments described in Section 2.2.4 to hold symmetry or asymmetry information. This is also true for the moments used for classification of the SOTON TTs in Table 5.5. These moments have proved useful in terms of subject separation implying that they significantly differ between subjects, producing possible unique symmetries for each subject. However, the remaining velocity moments which are directly comparable to those described in Section 2.2.4 vary between subjects. Assuming they encode symmetry information then these moments could provide more general symmetry information, which may be exploitable to detect human, or bipedal movement. Although, with reference to Equation 2.34, non-zero values are more likely due to the gait cycle not being exactly symmetrical, the summing over many images and the discrete implementation. For example the values for vm_{1100} have been observed to oscillate either side of zero suggesting possible symmetries in either axis.

These results support the hypothesis that symmetry information within the gait cycle, captured through the Cartesian basis (via the velocity moments) enables classification through the differences in a subject's symmetry of motion. In this

manner the velocity moments are describing symmetry (or asymmetry) within the different image sequences. With this in mind, the velocity moments of the silhouette data contain information about spatial symmetry within the image sequence. In contrast, when applied to the optical flow data they contain information describing the symmetry of temporal changes.

Murray showed that the amount of upper body sway was greater for males than females [55]. When viewed from the side this movement is visible as a vertical motion or 'bobbing' as the subject walks [13]. The Cartesian velocity moments used for classification of the SOTON TTs and the Zernike velocity moments applied to the CMU databases exploit this vertical motion, suggesting future possibilities for gender recognition. This also reflects the conclusion that the optical flow images (TTs) contain the inter-subject differences in joint and torso movements. (An alternative gender classification approach could simply exploit the differences in height between subjects. However, this assumes a fixed camera geometry and distance from subject. Also, the Zernike velocity moments re-scale the subject silhouette (to overcome variations in scale) removing any available relative height information. The results gained from the SOTON database demonstrate the availability of velocity moments that are invariant to direction of movement. The CMU database results emphasize that the features may be speed independent, provided that the gait cycle is not under-sampled or aliased. It must be noted though that a human running gait is fundamentally different to that of walking [84]. The two case studies have highlighted the effects of using different Zernike velocity moments, suggesting that different motions of a single subject are detectable, further emphasising that the velocity moments encode both shape and motion.

5.4.3 Summary of results

Two methods of velocity moments (Cartesian and Zernike), based on the same structure have been applied to seven different human gait databases, producing encouraging classification results. Table 5.27 collates all the classification results for the seven different databases. The final result separates the majority (96.00 %) of a 50 subject dataset using just twelve features. If there are X independent subjects then the probability of randomly correctly classifying each subject is:

$$C_s = \frac{1}{X} \tag{5.1}$$

Assuming that there are W independent sequences (samples) of each subject, and X independent subjects, then the probability of randomly correctly classifying one sequence (for one subject) is:

$$C_{ss} = \frac{1}{X^W} \tag{5.2}$$

Database	No. of Subjects	No. of gait cycles	Image dimensions	Velocity moment	No. of moments		Classification $k = 3$			Prob. of chance classifier C_{ss}
					STs	TTs	STs	TTs	STs+TTs	
SOTON	4	1	128×288 (w)	Cartesian	3	3	100.00 %	100.00 %	–	0.0039
UCSD	6	1	128×160 (w)	Cartesian	2	3	80.95 %	57.14 %	76.19 %	0.0036×10^{-3}
UCSD	6	1	128×160 (w)	Zernike	5	2	100.00 %	97.62 %	100.00 %	0.0036×10^{-3}
CMU_03_7_s	25	1	486×640	Zernike	6	–	90.00 %	–	–	0.04
CMU_03_7_f	25	1	486×640	Zernike	6	–	87.00 %	–	–	0.04
CMU_05_7_s	25	1	486×640	Zernike	6	–	92.00 %	–	–	0.04
CMU_05_7_f	25	1	486×640	Zernike	6	–	87.00 %	–	–	0.04
HiD	50	1.5	690×400	Zernike	8	5	83.50 %	50.50 %	96.00 %	0.00016×10^{-3}

Table 5.27: Comparison of the different database classification results ($k = 3$) and resolutions, where 'w' indicates windowed data.

and the probability of randomly correctly classifying the complete database for all independent sequences W, for each subject X is given by:

$$C_{ssd} = \frac{1}{X^{WX}} \tag{5.3}$$

The corresponding values of C_{ss} for each database are displayed in Table 5.27. The CMU databases consisted of four sequences per subject, each split from one long sequence, thus its value of C_{ss} is higher than that of the other databases. All of the chance probabilities are very low, and the values for C_{ssd} for each database would be essentially zero. The ability for the Zernike velocity moments to handle larger databases (reflected in the UCSD results for both the Cartesian and Zernike cases) reinforces the less correlated nature of the Zernike descriptors. These results are by analysis of relatively small databases, in comparison with other biometric databases, although, currently the HiD database is the largest of its kind. Possible specific velocity moments for gait recognition can only be determined through the analysis of larger databases. The results and analyses thus far have helped to isolate those moments more suitable to the classification of human gait. This is illustrated in both the CMU and HiD analyses where the list of possible moments were reduced using knowledge gained from the UCSD and SOTON analyses. Alternative feature extraction (i.e. subject extraction and optical flow) techniques exist, based around more complicated and sometimes intelligent methods i.e. [24]. Therefore the ideas, results and conclusions presented here are not solely dependent on the simple extraction methods used i.e. the background subtraction techniques used for the SOTON and UCSD databases. The ANOVA analysis has provided a simple way of reducing the feature set, however, this technique may fail for larger databases, not least due to the assumption that the samples of each subject (here 4) are normally distributed. The STs and TTs appear to complement each other, supporting the hypothesis that buttressing biometric techniques together is a valid avenue for further research. Increasing the number of sequences per subject may be desirable to improve clustering capability, also with respect to the k-nn classifier approach. A more intelligent classifier would invariably improve the results, aiding to avoid the problems of feature normalisation illustrated in the HiD analysis. In light of these considerations, classification results that encourage the use of gait as a biometric have been achieved, using simple feature extraction and classification techniques, and statistical features that encode shape and motion.

5.4.4 Comparison with other gait recognition studies

A direct comparison between different gait recognition techniques is not always possible. Such comparisons are hampered by differences between: the databases

that each technique is tested upon, the pre-processing or feature extraction techniques employed (i.e. subject extraction or optical flow techniques) and the final classifier techniques. As a result (in the majority of cases) comparison between results becomes a comparison of the complete methodology and not necessarily just the feature description technique. However, with this in mind, here follows a brief comparison of the results presented in this thesis with other current (at time of publication) gait recognition techniques.

The UCSD database analysis (Section 5.3.2) has improved upon the results by Little [46] on the same database. A result essentially due to the use of a scale independent descriptor (i.e. Zernike velocity moments), helping to overcome the problem of the varying distance between camera and subject - an issue not addressed by the original study. Huang [29] produces similarly high classification results for the UCSD database (using different silhouette data), although adding further subjects to this method of gait classification requires re-analysis of the complete database, unlike the method of velocity moments. The classification results for the CMU databases (Section 5.3.3) are higher than those achieved by Collins [11] and BenAbdelkader [5] using identical silhouette (ST) data. Higher classification results than those presented in Section 5.3.4 have been achieved on the Southampton ST HiD database by Hayfron-acquah [25] (the authors full results have yet to be published), however, these results have been achieved using a far greater number of features. Area masks, also applied to the ST HiD data [20] (the authors full results have yet to be published) produced lower classification rates than those presented in Section 5.3.4, however fewer features were available. An alternative method based on moments [41] (of single images, similar to the work by Little [46]) has achieved high classification results on a 24 subject database. The same technique achieved comparable results to those presented in Section 5.3.3 for the CMU ST databases. (However, this study [41] and other recent work [64] is more concerned with the problem of inter-subject variation caused by clothing, footwear, carrying object etc. An area which is beyond the scope of this thesis).

In conclusion, the results presented in this thesis have built upon previous gait studies producing improved or comparable classification results on identical databases. Similarly high classification rates have also been achieved on new larger databases, and further justification for the gait symmetry hypothesis has been presented. These high classification results have been achieved using relatively few features, while the number of features available are (in theory) infinite, and the addition of further subjects to the analysis is trivial.

Chapter 6

Performance analysis - Zernike velocity moments

6.1 Introduction

This chapter details the performance evaluation of the Zernike velocity moments as applied to the complete HiD ST database. The analysis is intended to provide an insight into the robustness of the technique under a selection of conditions, simulating possible application scenarios. This analysis has not been applied to the TTs as the results would be dependent on both the Zernike velocity moments and the optical flow technique. For example, we consider occlusion analysis. A partially occluded subject will produce different TTs. The results of analysing these new TTs will include secondary effects (effects of the occlusion on the optical flow calculation) rather than just characterising how the Zernike velocity moments perform under occlusion. In contrast, for the STs, the addition of occlusion or random noise is simple, due their binary nature. For each sequence of STs, the Zernike velocity moments used for classification (detailed in Table 5.24) were re-calculated for each increment step of the corresponding performance test (eg. for noise analysis, the velocity moments are calculated for each different level of image noise). The normalised mean squared error (NMSE) is then calculated between the original velocity moment values (O_i) (i.e. the noise free values) and the new 'altered' values (W_i), at each step of the performance test. The NMSE is defined as:

$$\text{NMSE} = \frac{\sum_{i=1}^{K}(O_i - W_i)^2}{\sum_{i=1}^{K} O_i^2} \tag{6.1}$$

where K is the number of features, or moments and a NMSE value of 1 indicates 100% variation from the original feature values. Describing the performance characteristics in terms of an error-rate produces an analysis that is independent of the characteristics of the database. For example, if this analysis was conducted using

the classification rate (instead of the NMSE), the results may be dependent on subject cluster compactness and separation. Additionally, the subject clusters may all shift in the feature space relative to each other (eg. as the amount of occlusion is varied), representing no change in the classification rate, even though the features themselves have changed. For each performance evaluation (except Section 6.4) the NMSE results are plotted for the complete HiD database, along with an example set of results for one randomly selected HiD subject. The results for the complete database display the NMSE mean (μ) and standard deviation (σ) relative to the mean ($\mu \pm \sigma$), indicating how the moment values disperse as the performance increment (i.e. noise) varies.

This chapter is structured as follows: The occlusion and image noise analyses (Sections 6.2, 6.3 and 6.4) are concerned with problems caused by the scene itself (i.e. static occluding objects and noise due to cluttered scenes or poor extraction techniques). The remaining evaluations: image resolution reduction and time-lapse imagery (Sections 6.5 and 6.6) are concerned with possible problems due to variations in image-acquisition hardware.

6.2 Occlusion

This analysis simulates to some extent the effects of a subject walking behind a lamp-post or another such static occluding object. It is interesting to note that human gait is self occluding, due the pendular arm and leg motion. A stationary occluding object can have one of two effects on the extracted subject ST. The first adds itself to the ST, the second removes a proportion of the shape. For example, both images in Figure 6.1 could be caused by a lamp post. In part, the differing effects will be due to the extraction technique, other factors include the camera viewpoint and also the item which is causing the occlusion. The performance of traditional moments degrades where the shape is occluded due to the loss in region-information. This is, in part, due to the moments being calculated from a single image. They are a global descriptor, so if a portion of the object is missing (or has increased in size), it does not seem unreasonable to expect the result to be different from that of the original un-occluded object. Depending on the size of

(a) (b)

Figure 6.1: A subject walking past a lamp post - two differing views of occlusion.

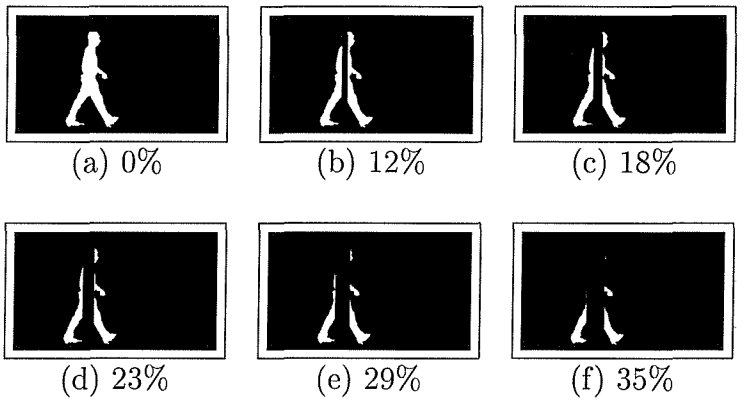
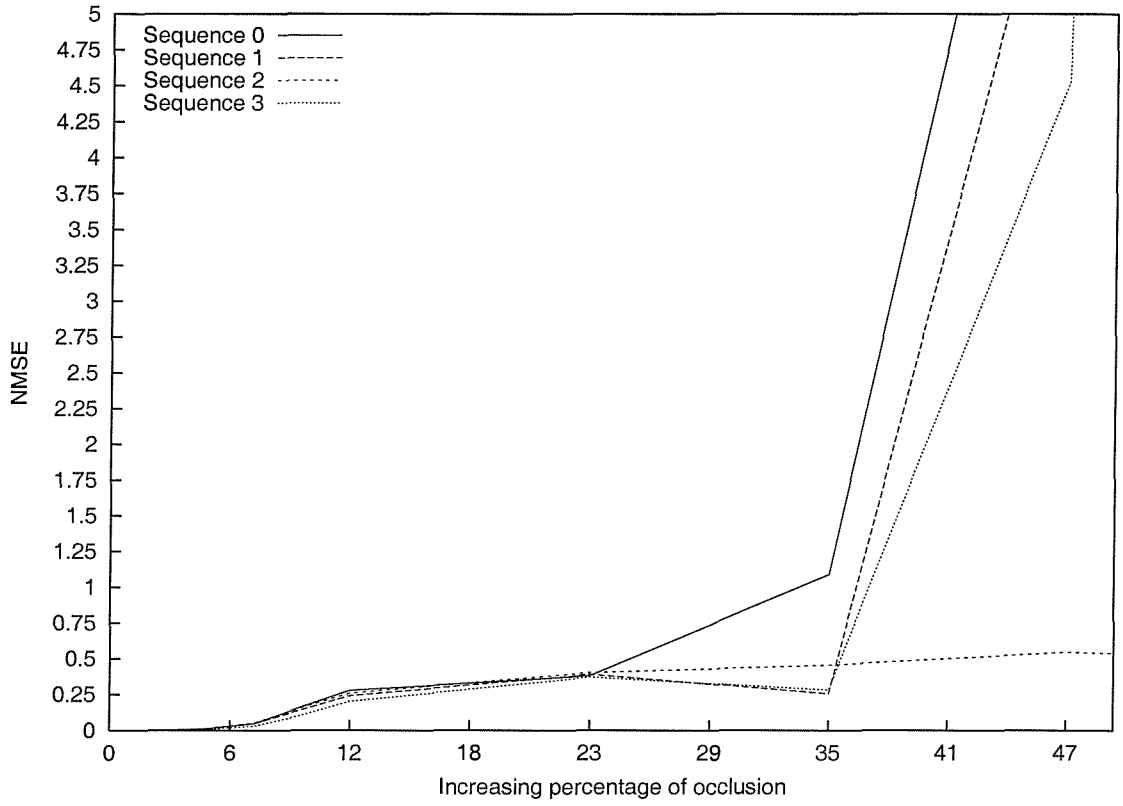


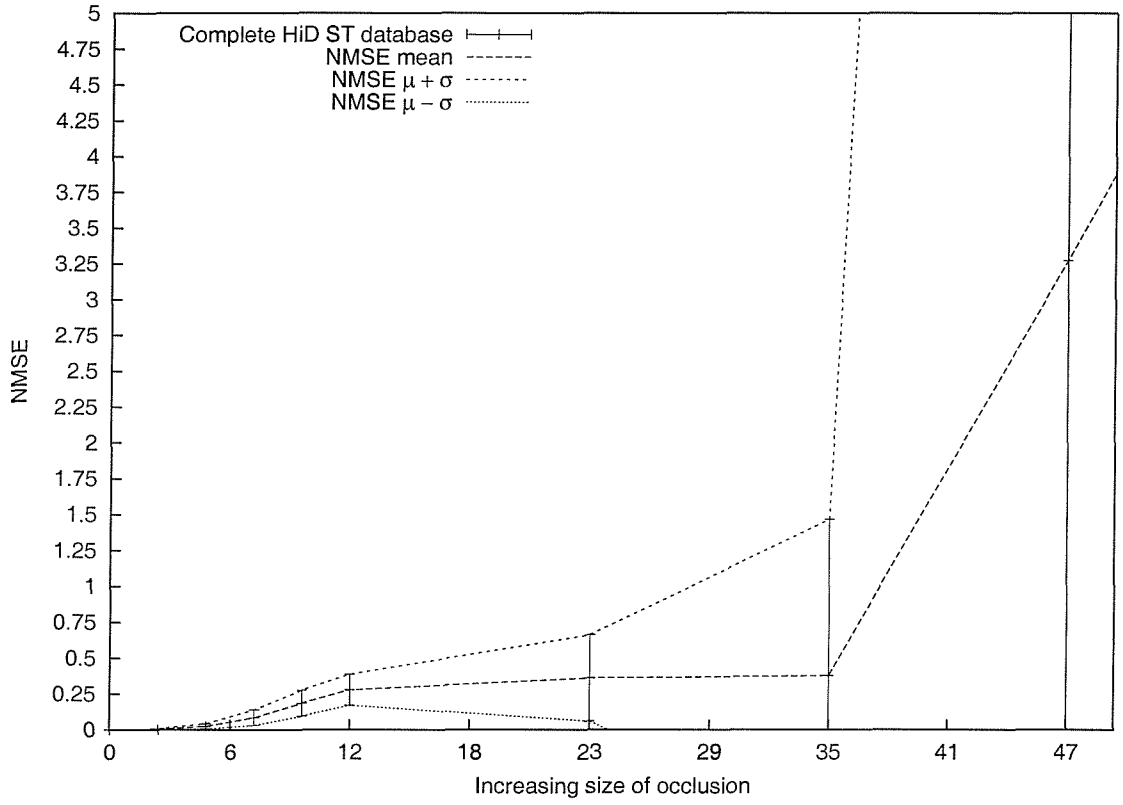
Figure 6.2: STs with increasing amounts of occlusion (percentage of gait cycle occluded).

the occlusion, even the lowest order moments (i.e. mass) will be altered. With respect to the Zernike velocity moments, adding an occluding object to the subject silhouette (Figure 6.1a) will cause the calculations to degrade rapidly. The addition of the occluding strip will bias the mapping function. Further, the calculation of the Zernike polynomials will be more efficient towards the edge of the unit disc (Section 2.4.2), favouring the top and bottom of the occluding strip, rather than the exterior of the un-occluded target object, i.e. the subject. For these reasons the analysis of the second type of occlusion, removing part of the ST (Figure 6.1b) is studied here. Figure 6.4 shows an example ST sequence of a subject walking through a stationary occluding strip. The Zernike velocity moments used for classification were re-calculated for the complete HiD database, at each occlusion increment. For each increment the NMSE, between the original un-occluded and the occluded moments was then calculated. The increment was determined in pixels, expressed here as a proportion of the average distance over which the subjects walked.

Figure 6.3a shows the results for one subject (4 sequences), whereas the results from the complete HiD database can be seen in Figure 6.3b. The NMSE is below 0.1 with 6% occlusion applied. The descriptions drift as the occlusion increases. The descriptions can be seen to become noisy and diverge ($\mu \mp \sigma$ increases) as the occlusion increases past 18%, which Figure 6.2 shows to occlude a large proportion of the ST. It must be noted that only one gait cycle has been used for the calculation. If, however, more than one gait cycle is analysed then the effects of the occlusion will essentially be further diluted, due to an increase in the spatial resolution, potentially providing more information about different parts of the gait cycle. Thus increasing the amount of occlusion that can be handled before the descriptions become overrun by noise. Similar effects will be true for all of the performance analyses.



(a) Increasing occlusion for one subject.



(b) Increasing occlusion for the complete HiD database.

Figure 6.3: NMSE with increasing occlusion for (a) one subject and (b) the complete HiD database.

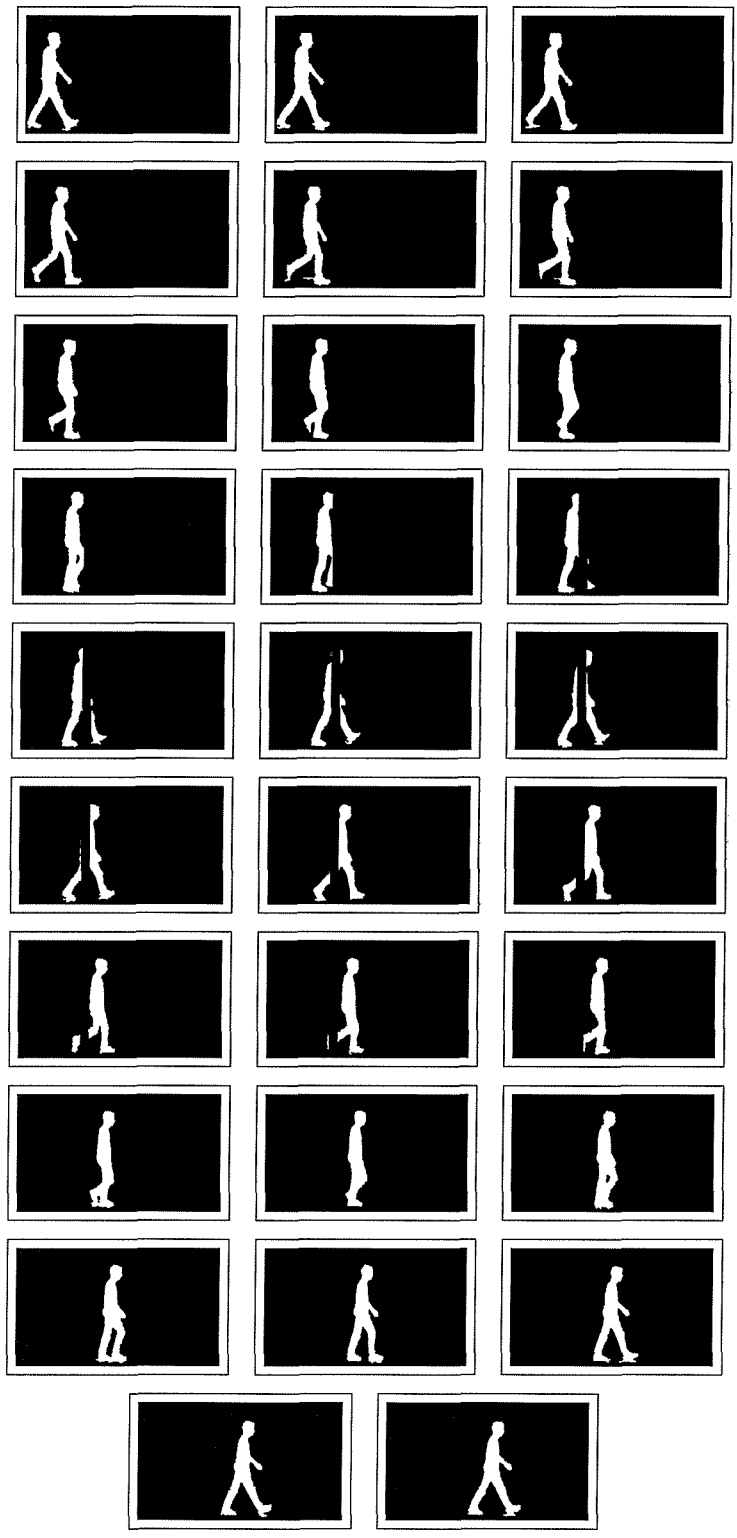


Figure 6.4: A sequence of STs showing the 18 % occlusion case. The subject is walking left to right and the sequence runs from the top left to bottom right.

```
seed uniform random number generator
    for (x → (width × height) )
        {
        noise_uni1 = uniform real random number from 0 to 100;
        noise_uni2 = uniform real random number from 0 to 255;
        if (noise_uni1 < amount)
            {
            if (noise_uni2 > 127) pixel[x] = 1;
            else pixel[x] = 0;
            }
        }
```

Figure 6.5: The pseudo code algorithm for the artificial noise analysis.

6.3 Simulated image noise

The analysis of the effects of simulated noise within image processing can be perceived as being very artificial, in the sense that noise can appear under many different guises. Noise within the system could be as a result of many different factors, take many different forms, with different noise distributions. Noise can be introduced into the overall system through both the hardware configuration (i.e. sensor noise in the camera pixel array, connecting leads - an effective antenna to background electrical noise, or even variations in mains supply or subject illumination), and any pre-processing on the raw data (i.e. extraction techniques, conversions between image formats, compression algorithms etc). Here we are attempting to illustrate the effects of general image noise on the Zernike velocity moments, to give an idea of their performance attributes where the data is randomly perturbed from its true value. Random uniformly distributed thresholded noise was uniformly applied to each image pixel in each sequence. The amount of noise varied from 0% to 100%, in 10% steps. For clarity, the pseudo code for the single image noise function (255 grey levels) is shown in Figure 6.5, where *amount* sets the percentage of noise to be added to each image. The use of the uniform (linear) deviate ($noise_{uni_1}$) to determine whether the noise is added allows a linear increase in the amount of noise added. The uniformly distributed deviate (between 0 and 255) replacing the pixel value ($noise_{uni_2}$) is thresholded at 127 to provide binary noise, as the STs are binary. Thus a Gaussian distribution (thresholded in the same manner) would produce a comparable result (with $\mu = 127$ for 255 grey-level image, ignoring any rounding effects at the mean), however, using the uniform distribution implementation is computationally simpler. Figure 6.6 illustrates three images from the HiD database, shown with increasing amounts of noise. The effects of applying the noise before and after the mapping process (Equation 2.58) are considered. This is

done due to the degrading effects that the noise can have on the mapping function. However, we are primarily interested in the effects of the noise on the Zernike velocity moments. Considering first the before-mapping case. The mapping process depends directly upon mass and centroid information. The random noise quickly degrades this information, causing a 'domino' effect. The noise degrades the mapping function, while the mapping function in return effectively amplifies the noise, the results of which are then passed on to the Zernike polynomials. The final calculations degrade rapidly from their true value, as can be seen in Figure 6.7, shown for one sequence from the HiD database. Applying the noise to the complete image causes the mapping function to map the complete image (rather than just the subject silhouette) into the unit disc. This occurs as soon as any noise is introduced. As the noise increases the mapped silhouette becomes smaller (with respect to the area covered by the noise). This is due to the mapping process maintaining the image mass for scale invariance. The Zernike polynomials evidently describe the noise with constant image mass. Hence the noise error rate increases rapidly, as seen in Figure 6.7. The analysis was not repeated for the complete HiD database due to these initial results and conclusions.

However, applying the noise after the mapping process produces the results shown in Figure 6.8a. The NMSE for the complete HiD database steadily increases with the noise level. Here, as before, the error bars show the standard deviation σ (with respect to the mean μ) of the moment values for each level of noise. The standard deviation can be seen to be very small. Figure 6.8b shows a similar plot to that of Figure 6.8a, with $(\mu \mp \sigma)$ replaced by the (NMSE) minimum and maximum values (about the mean value) illustrating that some variation about the mean does occur. These results reflect the effect of the subject silhouette being spread, or diluted, across the unit disc as the noise increases. As the noise increases the subject silhouette has less and less of an effect. The noise is applied to the complete unit disc and noise pixels appearing close to the unit disc's perimeter will have more weighting effect on the moment values than the silhouette (which is located about the origin - centre of the unit disc). The random nature of the images being encoded will in turn produce random Zernike moment results (for each image), the values of which oscillate over the possible moment value range i.e. positive and negative about zero. Therefore the summed (or average) result of each image's Zernike moment (i.e. the velocity moments) of a sequence will approach zero as the noise and sequence length increase. As the velocity moment values approach zero, so will W_i in Equation 6.1, hence, as the noise increases the NMSE will degrade to 1.

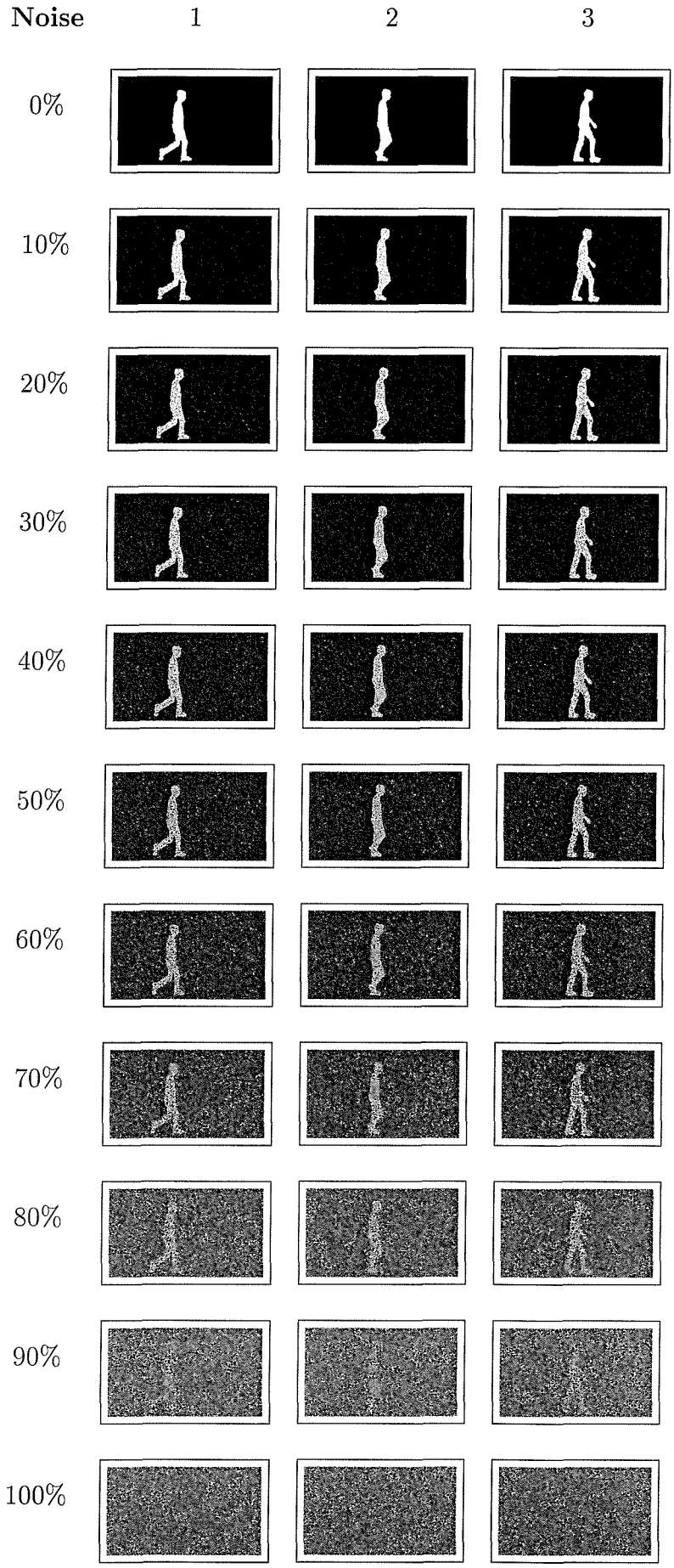


Figure 6.6: Part of a subject sequence showing increasing amounts of noise.

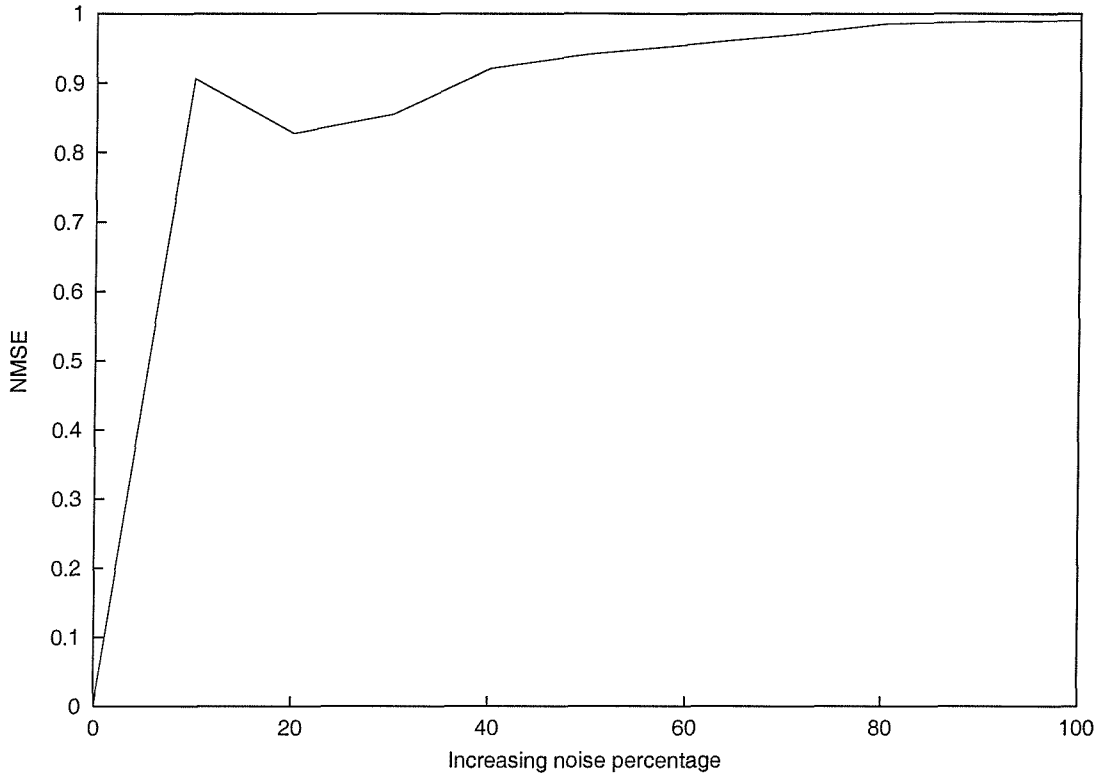
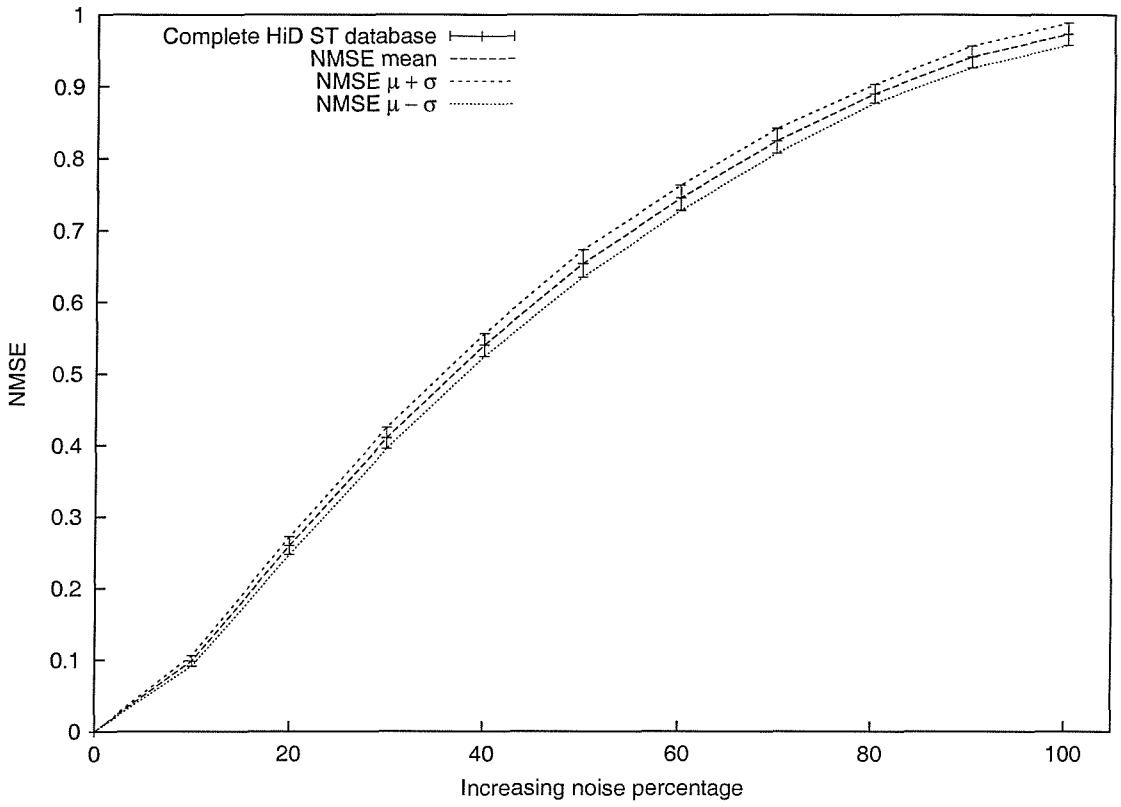


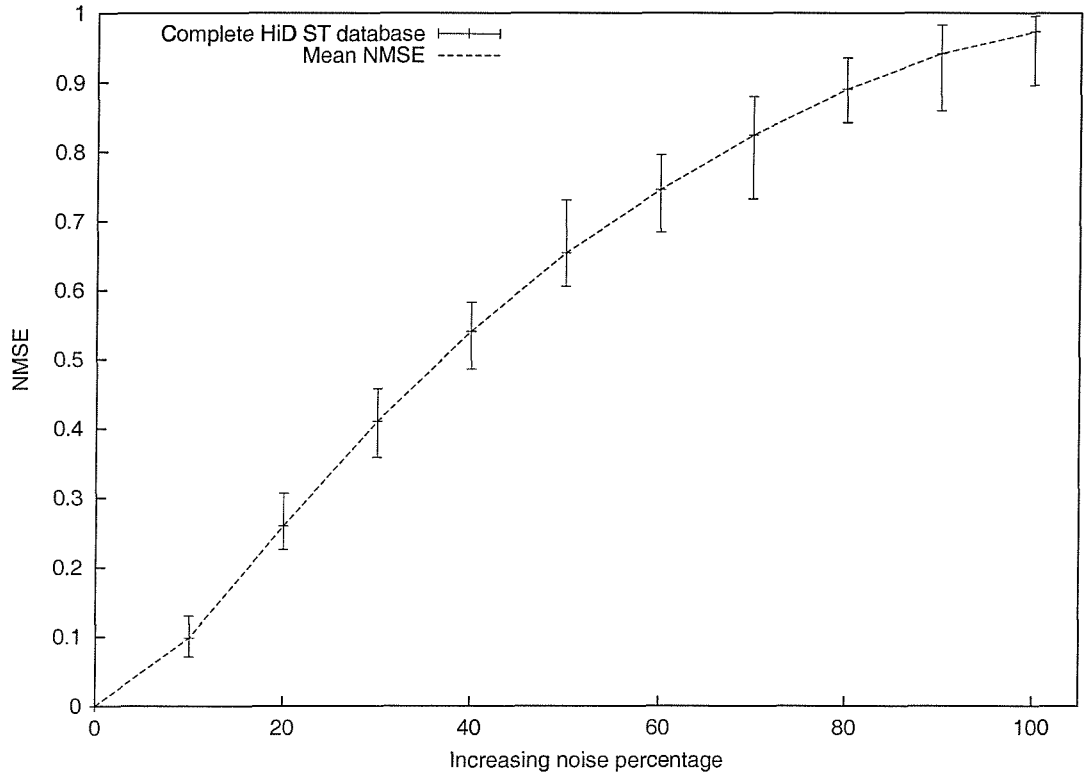
Figure 6.7: NMSE with increasing noise (applied before mapping) for one sequence from the HiD database.

6.4 Real-world image noise

Due to the artificial nature of the previous study on noise, an analysis based on real-world noisy data was carried out. This analysis is aimed at studying the translation of the velocity moments from the laboratory data, to data that is less controlled (i.e. no controlled lighting, pedestrian and vehicle scene noise etc). However, the results will essentially analyse the performance of the extraction technique. An in-depth analysis on the whole HiD database would yield results which were highly dependent on aspects like the meteorological conditions and how they affected the extraction. Thus, we have chosen to study only a small part of the HiD database in an attempt to illustrate possible affects of using real-world data. This outside data consisted of two subjects from the HiD (indoor) database, with four sequences of each subject walking for three consecutive heel strikes (as per the HiD indoor database). The images were captured outside on the same day as the indoor data comprising the HiD database, so the subjects have the same attire. Example extracted images can be seen in Figure 6.9. The background scene is visually noisy (i.e. cars, pedestrians, bicycles, etc) thus affecting the subject extraction. Holes in the ST and shadows on the ground are apparent in Figure 6.9b, whereas part of a car (which happens to be moving at the same speed as the subject) has been extracted in Figure 6.9a. The subjects were extracted using a colour implementation



(a) NMSE plotted with mean and standard deviation estimates.



(b) NMSE plotted with mean, minimum and maximum values.

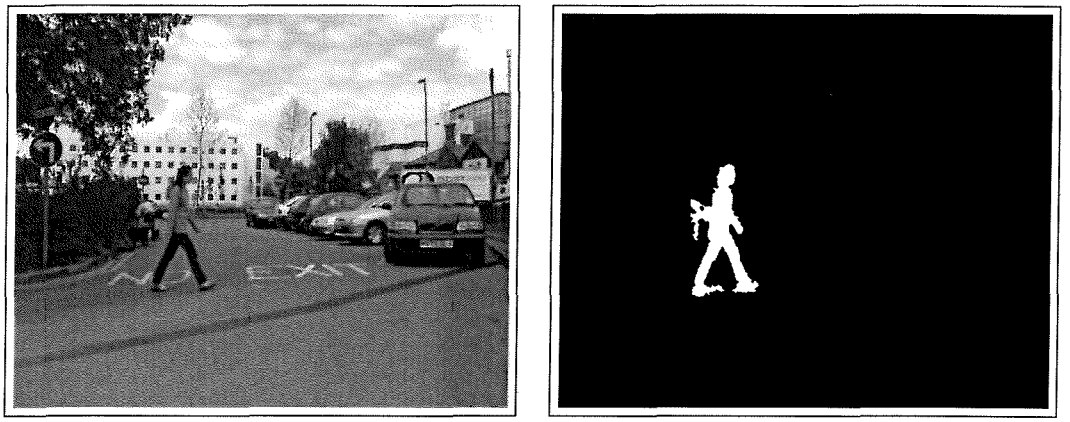
Figure 6.8: NMSE with increasing noise (applied after mapping) for the complete HiD database.

of the statistical extraction technique [33], described in Section 4.4.2 and Appendix B. As the subject's trajectory was known, the area outside of this was masked - this helped to reduce the scene noise, while also speeding up the (computationally expensive) extraction. Windowing the subject may further improve the extraction results - this could be achieved by automatic techniques, i.e. [57, 24].

The selected Zernike velocity moments used for ST classification were re-calculated on this outside data. The HiD ST database (captured inside and extracted using chroma-keying) was used as a benchmark or ground-truth, and estimates of the within-class (or subject) variances along with the cluster centroid movement (between inside and outside data) are presented. The moments were first normalised to the maximum values across all sixteen sequences (four per subject, two subjects and two sources of data - inside and outside) removing any possible bias from moments with naturally large values, while allowing between subject comparisons. The shift in each outside data cluster centroid is expressed as a Euclidean distance, measured from the corresponding inside data cluster centroid. The multidimensional within-class variance for K samples, using the Euclidean distance metric d (Equation 4.14) is expressed as:

$$\sigma^2 = \frac{1}{K}\sum_{j=1}^{K} d^2 \tag{6.2}$$

Table 6.1 shows these results and Figure 6.10 displays the results graphically, where each cluster centroid μ and variance σ^2 is represented by a Gaussian distribution. Both subject cluster centroids (outside data) μ have drifted from their original (inside data) value, whereas the within-class variance σ^2 for subject 012 has increased dramatically. This is most likely due to the large shadows (Figure 6.9b) apparent throughout the four image sequences (caused by bright sunlight), a characteristic which is not present on the inside data. Subject 037's within-class variance has also increased, again reflecting the addition of background noise and slight shadows around the feet appearing in the ST (Figure 6.9a). These additional objects appear and disappear through the sequences. This can be seen in the example outside data sequences (original and STs) in Appendix D. (In contrast, the holes present within the subject silhouettes in the CMU data (Section 5.3.3) are consistent throughout the image sequences, effectively increasing the amount of usable subject spatial information). However, it must be noted that these results are only a reflection of how these particular velocity moments (jointly) behave under less favourable scene conditions. Specific moments will invariably be affected by different changes in image content. Possible moment selection could be based around both the ANOVA analysis of clean (inside) data and analysis of outside data, utilising moments which exhibit good inter-subject separation and reduced variation



(a) Subject 037

(b) Subject 012

Figure 6.9: Example images from the outside data, along with their corresponding STs.

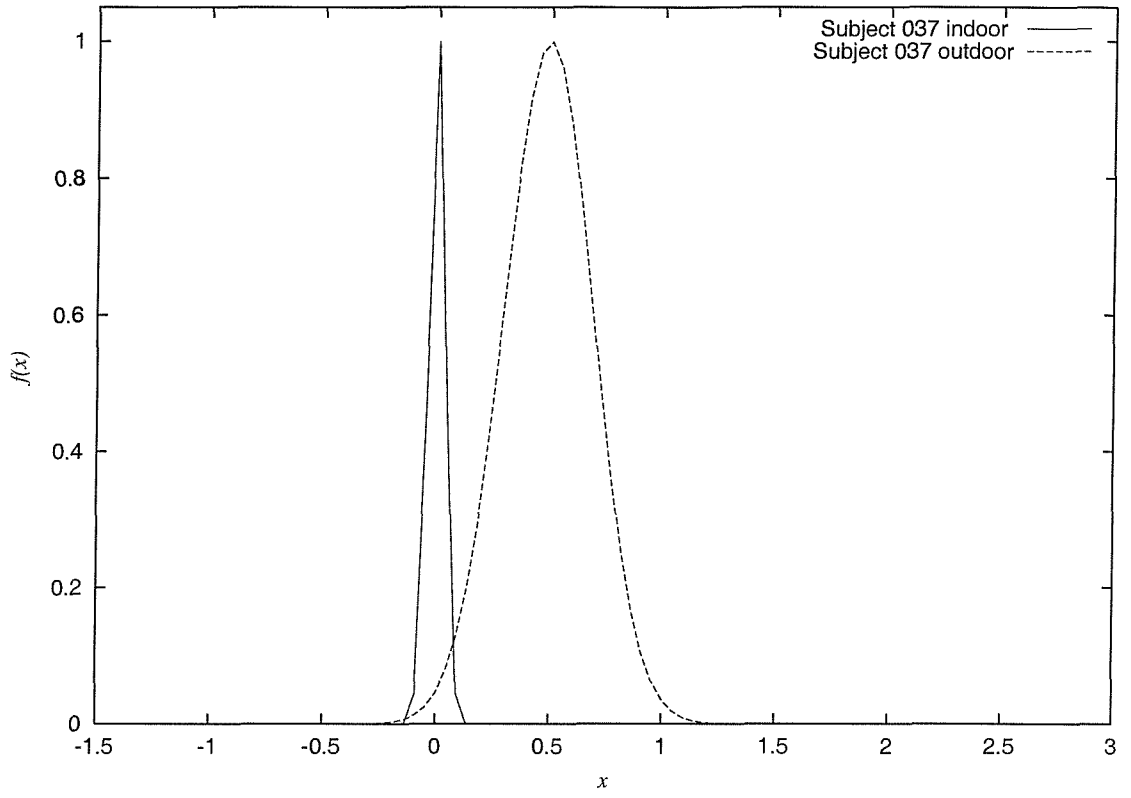
Subject	Inside σ^2	Outside σ^2	Outside μ shift
037	0.0013	0.0389	0.4906
012	0.0034	0.3764	0.8874

Table 6.1: Comparing the inside data velocity moments with those calculated on outside data.

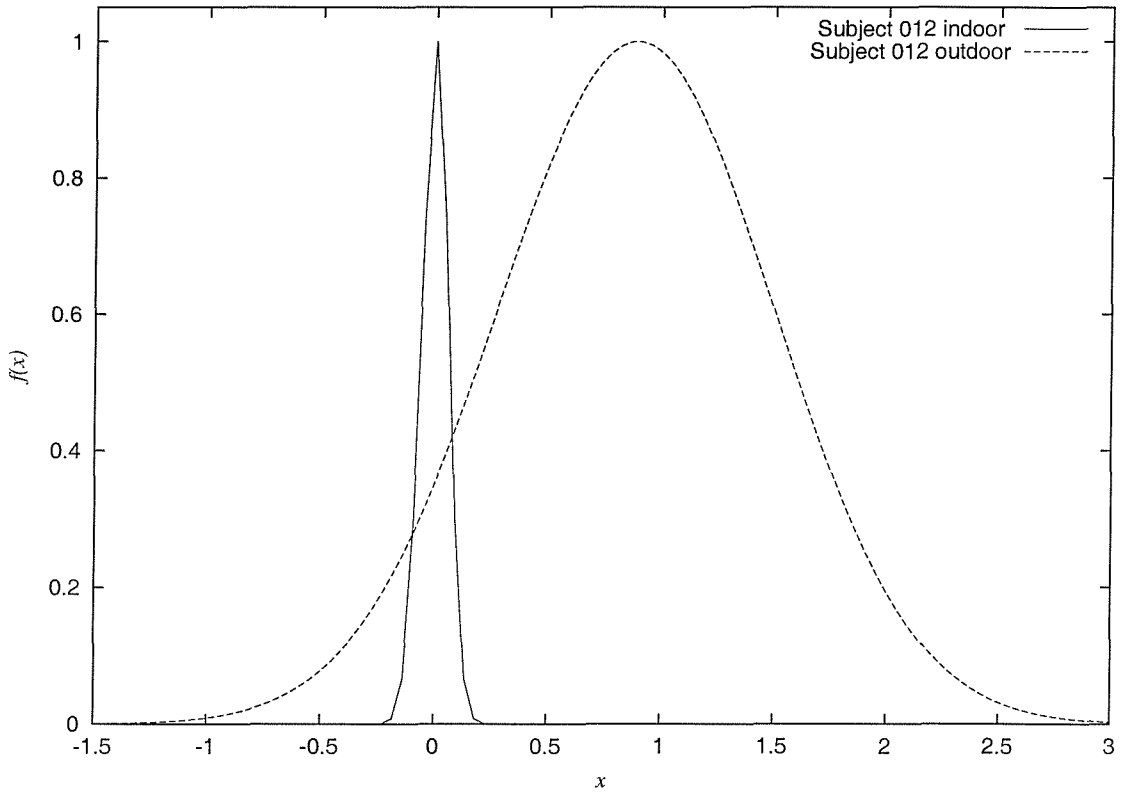
due to outside scene noise. Further, longer image sequences will invariably improve the results by exploiting sequence correlation.

6.5 Image resolution

Camera resolutions vary considerably between different manufacturers, while the distance from the camera to the point/area of interest will also vary, dependent on the application. Gait as a biometric has the unique advantage of being potentially detectable from a distance (unlike for example, iris or fingerprint analysis). By analysing the effects on the velocity moments of reducing the image resolution, an insight can be gained into how the technique may translate to lower resolution imagery. In this way, an idea can be gained of the minimum resolution needed before



(a) Subject 037



(b) Subject 012

Figure 6.10: Gaussian representation of the within-subject distributions for indoor and outdoor data.

the moment values diverge grossly from their original value, effectively becoming overrun by noise through loss of image information. If this reduction in resolution is performed before the mapping function then the results are dependent both on the mapping calculation and the Zernike moment calculation. However, the mapping process will have a positive effect on the handling of lower resolution images. It effectively ensures that there is no loss in the accuracy of the Zernike polynomial calculations, by mapping the reduced resolution image to the same grid size as the original resolution calculation, thus making the two results directly comparable. If the reduction in resolution is applied after the mapping process, theory suggests that the errors will rapidly increase due to loss of both image and calculation precision. Therefore, here we have studied the effects of applying the reduction in resolution before the mapping process. Assuming that the original image is the highest resolution available, the images were progressively re-sampled to reduce their resolution. Sub-pixel estimation is allowed, enabling any re-sampling size to be achieved. A detailed description of the image re-sampling algorithm can be found in Appendix C. Eleven different resolutions were analysed from $\frac{1}{2}$ the original resolution, through to $\frac{1}{50}$. At each different resolution, the previously selected Zernike velocity moments used for classification were calculated. The NMSEs were then calculated between the original resolution velocity moments and the reduced resolution versions. Figure 6.11 shows an original ST and reduced resolution versions, shown both expanded to their original and their relative sizes. Whereas, Figure 6.12a shows the NMSE plotted as the image resolution decreases, shown for one subject (four sequences). The x axis is the relative pixel size n, where $\frac{1}{n}$ is the new resolution. Figure 6.12b shows the mean μ results for the complete database, with error bars indicating the standard deviation σ of the NMSE for each image resolution. It can be seen that the errors begin to diverge ($\mu \mp \sigma$ increases) as the pixel size increases past 10. However, the NMSE errors are still low, less than 0.02. A slight increase in variance can be seen at $n = 5$ in Figure 6.12b. This may be due to moving decision boundaries in the re-sampling algorithm causing an increased error rate for a selection of subjects, effectively a rounding error. There are two possible reasons for the overall low NMSE values shown in Figure 6.12b. The first is with reference to the selected velocity moments themselves, which give measures of average pixel distribution in both the x and y directions. These properties will steadily degrade, however they will still be present until just before the image becomes one large pixel, refer Figure 6.11. Further, even though the image sequence resolution is being degraded, the overall x velocity will stay relatively consistent, as this is calculated using the COMs. The second reason for the low NMSE is (as already mentioned) due to the mapping process. The re-sampled images are passed onto the Zernike velocity moments for calculation at the same unit disc resolution, while the data itself is

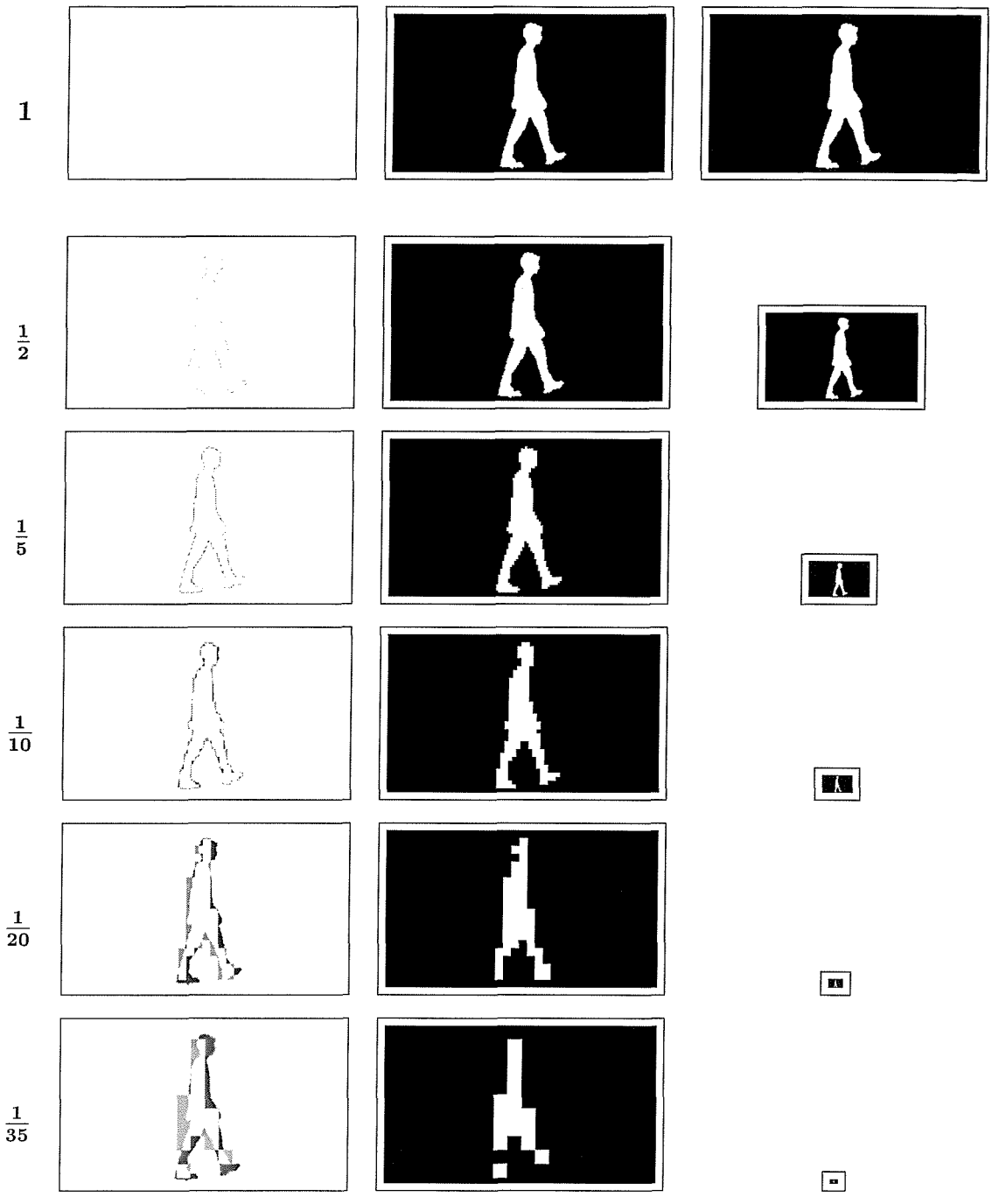
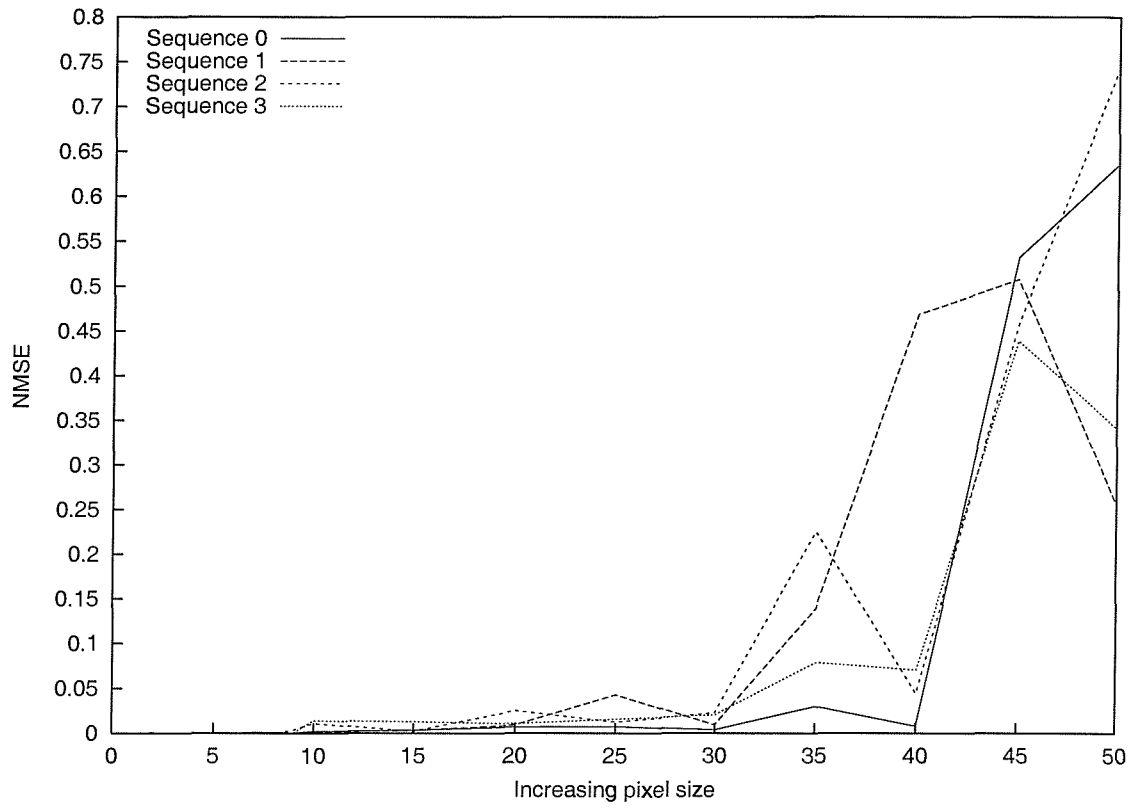
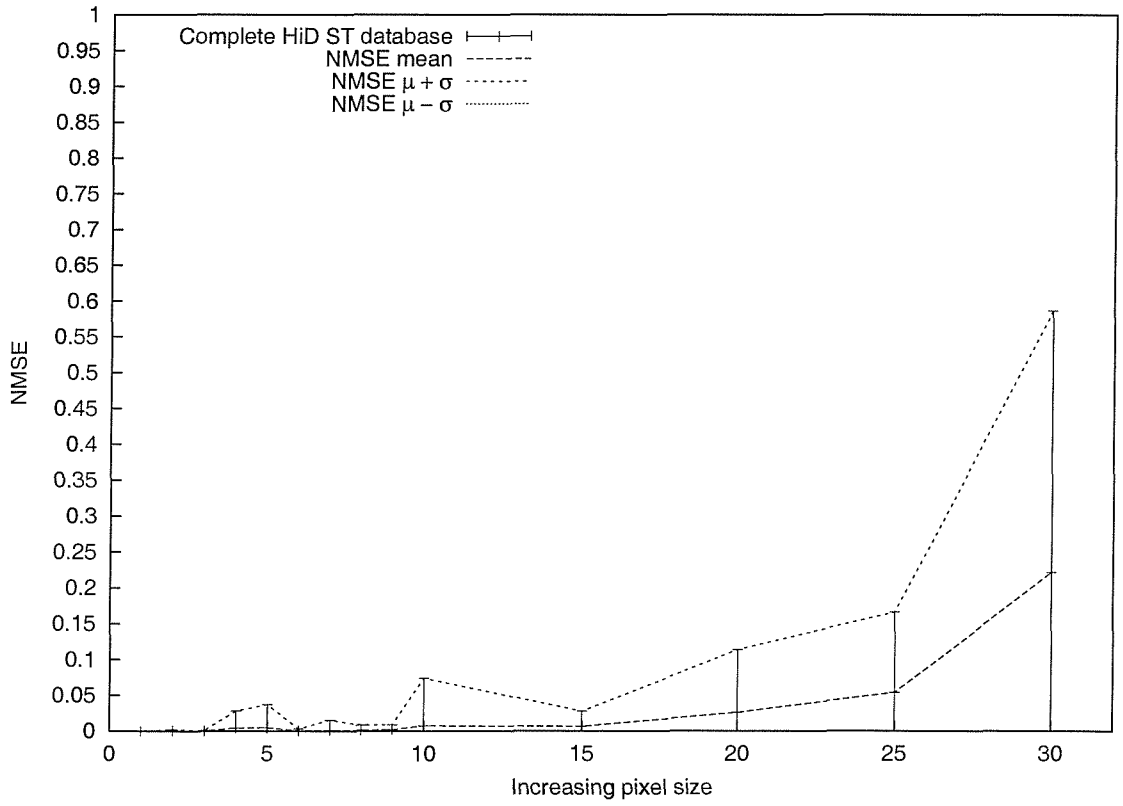


Figure 6.11: ST resolution degradation, original at the top, (showing from left to right) the re-sampling scalar, the difference image, resultant re-sampled image and their relative sizes.



(a) Decreasing resolution (increasing pixel size) for one subject from the HiD database.



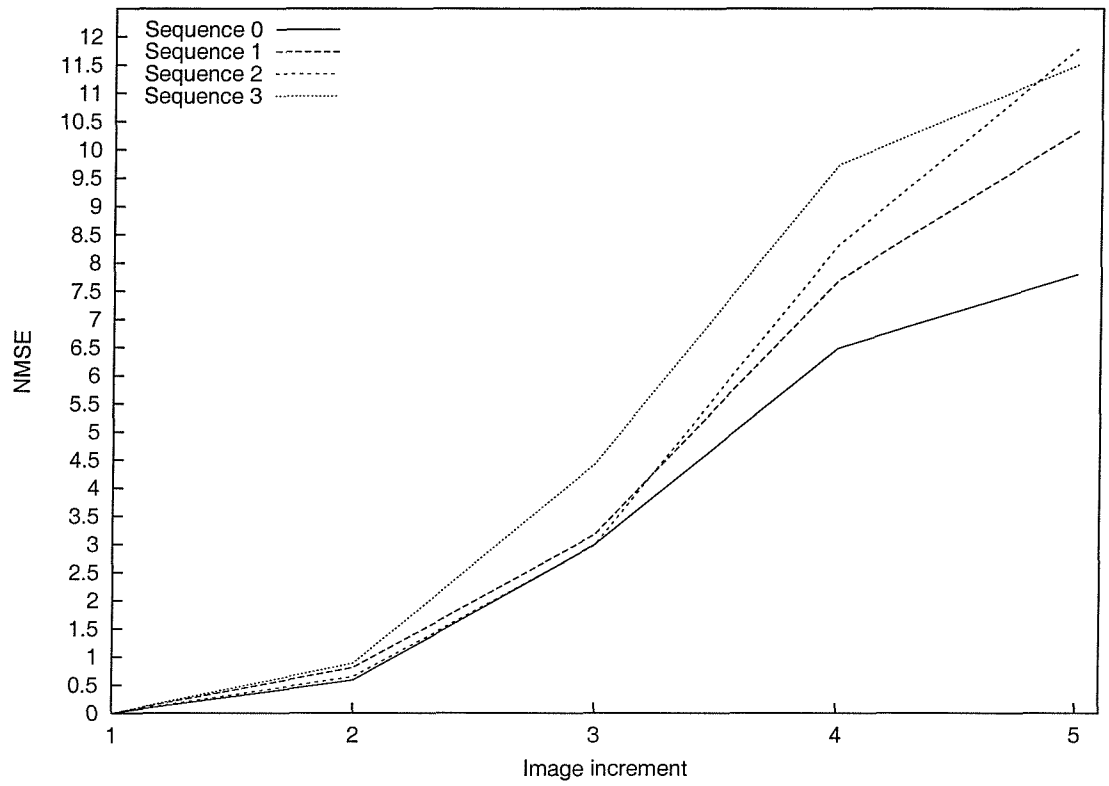
(b) Decreasing resolution (increasing pixel size) for the complete HiD database.

Figure 6.12: NMSE with decreasing resolution for (a) one subject and (b) the complete HiD database.

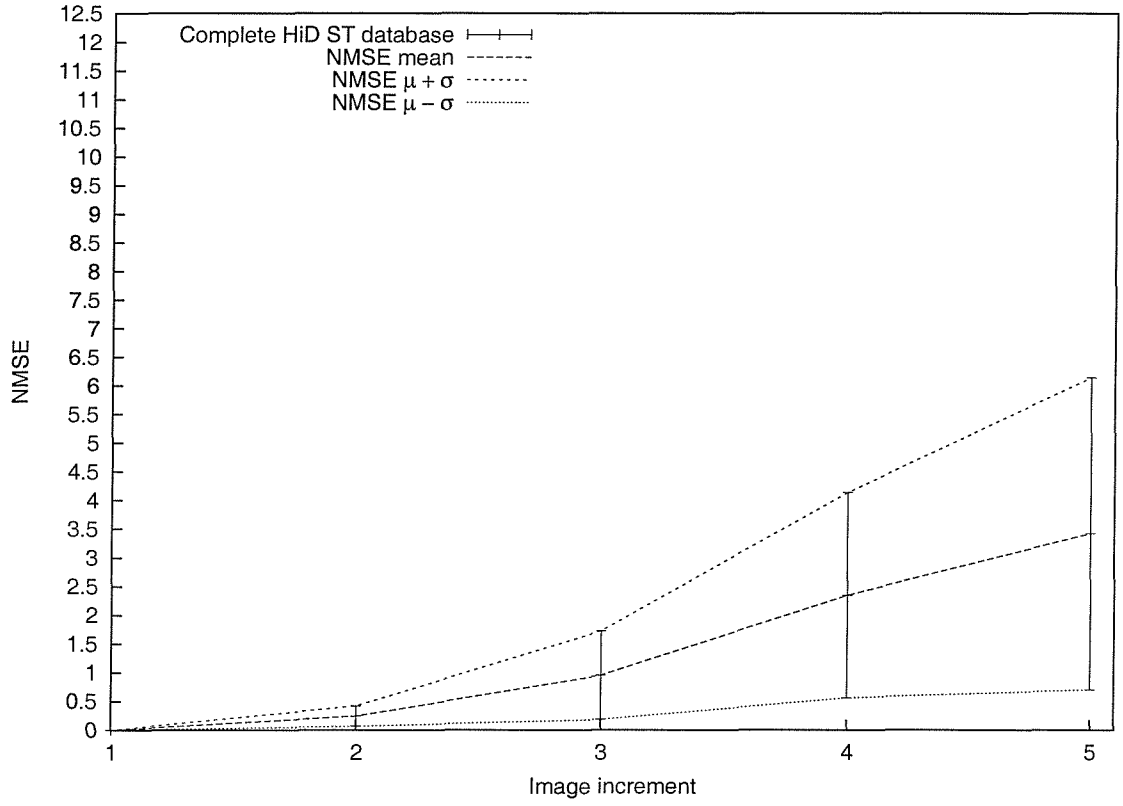
'grainier'. Thus, even though the image resolution has been reduced the accuracy of the calculation has not. Degrading the resolution is effectively adding noise to the perimeter of the silhouette, up to the point where each image loses its overall shape. This can be seen in Figure 6.11, where a re-sampled image has been overlaid onto the original resolution image, producing the difference images. The dark grey areas are the remains of the original silhouette after the differencing operation. The light grey areas are the remnants of the re-sampled image. It can be seen that pixels have both been added, and removed from the original silhouette by the re-sampling operation. As a final point, even at the $\frac{1}{20}$ resolution very little specific spatial information (except that from motion) will be available (as illustrated in Figure 6.11), whereas the NMSE error is still relatively low at < 0.05 (Figure 6.12b), reinforcing the advantage of using a correlated temporal image sequence.

6.6 Time-lapse imagery

This analysis is aimed to provide an insight into the effect on the Zernike velocity moments of reducing the temporal resolution. Images were successively removed from each sequence, to simulate the effect of time-lapse imagery, present in most low cost, restricted bandwidth surveillance camera systems. The analysis is restricted both by the frame rate of the original data, here 25 frames per second (fps), and by the length of gait cycle. The sequence length of one complete gait cycle is typically 30 images (viewed from a distance of $\simeq$ 3m and captured at 25 fps). Halving the frame rate considerably reduces the temporal resolution. Table 6.2 demonstrates the method used to reduce the frame rate (demonstrated for only nine images). To enable this analysis the Zernike velocity moment calculations use the scale, time and sequence length normalisation described in Section 3.5. As before, the velocity moments used for the classification process (for the complete HiD ST database) were re-calculated as increasing numbers of images from the ST sequence were removed. Figure 6.13a shows the NMSE results for one subject, while Figure 6.13b shows the results for the complete HiD database. As expected, reducing the temporal resolution can significantly affect the velocity moments. It must be noted that different combinations of velocity moments will produce different NMSE results, however, the trend of the results will remain the same. For example, velocity moments describing purely average x direction motion will be less affected by time-lapse imagery, than those describing purely structural information. The structural description will become increasingly diluted due to the increasing lack of specific spatial information. Figure 6.13b demonstrates this, as the frame rate decreases (image increment increases, see Table 6.2) the range of the NMSE increases rapidly. The increasing NMSE is occurring across the complete dataset, as shown by the increasing NMSE mean (μ) and standard deviation (σ) values.



(a) Decreasing frame rates for one subject from the HiD database.



(b) Decreasing frame rates for the complete HiD database.

Figure 6.13: NMSE with decreasing frame rates (increasing image increment) for (a) one subject and (b) the complete HiD ST database.

Image Increment	Images numbers present									Equivalent Frames/Second
	1	2	3	4	5	6	7	8	9	
1	X	X	X	X	X	X	X	X	X	25.00
2	X		X		X		X		X	12.50
3	X			X			X			8.33
4	X				X				X	6.25
5	X					X				5.00

Table 6.2: The equivalent time lapse frame rates, showing the construction of the image sequences from the original 25 fps, shown here for only 9 images.

6.7 Discussion

These performance analyses were designed to provide an insight into the characteristics of the velocity moment framework, and more specifically the Zernike velocity moments. However, performance metrics which involve artificially created scenarios will invariably produce artificial results. They can however be used to gain an insight into how these scenarios may affect the technique under test, highlighting possible advantages or pitfalls. For instance, the results of any noise analysis on a system will be dependent on the noise model employed. Here we are analysing statistical moments, which are used to describe the distribution of an image plane. Thus, a different noise model will alter the distribution in a different manner, producing different results. However, the overall conclusions should remain consistent. Where the noise model is not only the type of noise distribution employed (uniform, Gaussian, Rician etc), but also includes the way in which the noise is applied to each pixel and the method used to vary the amount of noise. Whether the noise be additive, replicative and/or varied by altering the noise variance, or dependent on a further distribution. One further consideration is the point of entry of the specified noise (i.e. scene noise from lighting effects, camera sensor noise, background electrical noise affecting the connecting cables etc). This in turn may determine the distribution and/or the model used. Here we have used a Gaussian distribution (using the assumption of the central limit theorem).

In terms of performance analysis, first we have looked at the problem of perimeter noise, simulating poor extraction of a contour (Section 3.2.3), then moving on to scene 'salt and pepper' noise in Section 6.3 simulating possible camera sensor and/or noise produced through transmission or image compression techniques. The 'real-world' noise analysis has highlighted the possible need for velocity moment selection based both on ideal data and application based data. We have already mentioned the different ways of perceiving occlusion (Figure 6.1), the result of which is primarily determined by the feature extraction technique. The image resolution analysis attempts to address the problem of increased distance between camera and subject, zooming and the effects of reduced resolution images - often

incurred through low cost surveillance systems or home movie cameras. However, this analysis is very much dependent both on the original image resolution which is re-sampled (used as the ground truth) and the re-sampling algorithm itself. Further to this there are also lens effects - smaller resolution cameras will invariably use lenses with different radial effects, a variable which is dependent on both the quality of the camera and its physical size. Further, zoomed imagery will be affected by radial distortion in a different manner to scaled imagery. The time-lapse analysis has highlighted the effects of reduced temporal resolution, the results of which are very much dependent on the number of gait cycles analysed. This is also true for all of the performance analyses. Longer image sequences will effectively dilute the effects of noise, occlusion and reductions in both spatial and temporal resolutions, by increasing the temporal correlation. Overall for the data studied here, the velocity moments appear least sensitive to the image resolution and most sensitive to the temporal sampling. However, the results will always be highly dependent on the feature segmentation. The image calibration issues discussed in Section 5.4.1 are equally applicable to these performance analyses. Finally, it is interesting to note that the scale and translation invariant mapping can essentially be perceived as a pre-processing technique used to improve the results and properties of the Zernike polynomial calculations.

Chapter 7

Future work

Two approaches to further this research are discussed - technique and application. The first is concerned with furthering the theoretical moment theory and technique, with aspects including alternative feature analysis techniques, optimisation of the velocity moments, the effects of altering the properties of the images/templates being described and the possibility of altering the moments' basis function. The second includes reinforcing the gait studies carried out so far, by increasing the depth and variation of the subject database. This will help to clarify the results gained and to increase the understanding of the velocity moment metrics, i.e. which ones are most useful in human gait recognition. Aspects including variations in clothes, baggage and even footwear have yet to be investigated. This avenue of research includes gait-specific issues regarding the recognition of types of motion. Lastly, the analysis of alternative applications would doubtless aid further understanding of this new technique. Alternative moving-shape applications include medical imaging (analysis of injured joint movement as compared with healthy movement - i.e. vertebrae damage to the lower back), astronomy (observing comets as their shape degrades due to the effects of planets or collisions) or even biology (monitoring cell growth and movement within a solution).

7.1 Technique

7.1.1 Alternative feature analysis and test

One critique of the feature selection methods presented here is that the selected features can tend to be dataset dependent. This is in part due to the limited sizes of the databases (as mentioned in Section 5.4.3), and due to the differences between them i.e. outside data versus treadmill data and different spatial resolutions. For example, using a (selected) moment list from the UCSD database as potential selected features for the CMU or HiD database may prove inconclusive. The CMU database consists of treadmill data in comparison to the outside UCSD (non-treadmill) data. Whereas the HiD data will have improved vertical motion resolution, due to the

increased image spatial resolution, in comparison to the UCSD data which has less than half the spatial resolution. In light of these differences the possible moment list was gradually reduced to those which showed promising attributes for human gait classification, through the analysis of multiple databases. (The method used may be more attuned to promoting further gait analysis, as the databases studied to date will not be fully representative of a large population, suggesting the possibility that some gait components needed for separation of a larger sample are not present in this reduced moment list.) These selected features were not necessarily the optimum set, both due to the final manual intervention after the ANOVA analysis and the ANOVA analysis itself. The ANOVA technique suggests which features are useful in separating the dataset. As the dataset increases in size, the need for multiple features becomes more apparent. The single-way ANOVA selects features that singularly separate portions of the dataset. Whereas we are actually using multiple features to classify, suggesting an n-way ANOVA may prove more useful for larger datasets, enabling the analysis of the interaction between features for an n dimensional feature or classification space - a topic that is touched upon in Section 4.3. This more in-depth analysis will produce a massive computation overhead as producing an optimum solution involves testing all combinations of moments.

An alternative method of analysis of the selected features is possible by dividing each database into two. This would enable more independent selection - ideally implemented using a larger database. By training, or selecting features on one half of the dataset and then testing or probing the remainder using the selected features, the problem of dataset specific features can be addressed. This approach was used in a recent study [64] where they describe a gait challenge centred around a large 74 subject ($\approx$ 300 Gigabytes) database. With reference to the work presented in this thesis, it is noted that the velocity moments used for the classification of the CMU databases use identical moments for the two different speeds for each view (i.e. CMU_03_7_s versus CMU_03_7_f and CMU_05_7_s versus CMU_05_7_f), essentially training on one gait speed dataset, then testing on the second speed dataset, producing high classification rates for both cases - an example of dataset independent selection.

Finally, presenting the performance analyses in terms of classification rates may complement the techniques detailed in Chapter 6 when comparing (application based) classification rates, especially if the compared techniques have been applied to the same database.

7.1.2 Computational demands

The research to date has been concerned with development and analysis of the properties of the velocity moments. However, once a technique is applied, the need

for faster computation becomes apparent - the need to optimise existing algorithms. For example, speed increases can be easily obtained by the use of look-up tables for values which are used repeatedly eg. mapping functions from Cartesian to polar coordinates. Two main areas of optimisation are discussed here, the first concerns the method of the moment calculations, the second with hardware configuration. The representation of a digital image used here is a collection of pixels, with a intensity value for each pixel. In this representation the double integral of a 2-dimensional moment, Equation 2.13 is replaced by double summations, Equation 2.16. The direct computation of such an equation requires a large number of additions and multiplications, which are computationally demanding. Many studies [35, 44, 77, 82], have looked at the problem of making moment computations *real-time*. Two main approaches are apparent: the first is to exploit the properties of either the moment calculation itself or the images being compressed. For binary images, it is possible to describe the region of the shape using chain codes (or other boundary descriptors), and then compute the moments based on the assumption that the area within the boundary is filled. If the shape to be compressed is a polygon, the moments can be calculated using the vertices [77], or corner points of the triangles which make up the polygon [82]. An alternative approach is to calculate the line segment integrals [44], converting the calculation of two dimensional moments into a one dimensional problem. This method appears to be applicable to the Cartesian velocity moments.

Statistical moments are ideally suited to implementation via parallel computing. This involves altering the algorithm to exploit the independence of operations. In terms of the velocity moments, the calculations for each image sequence within a database are independent. Further, the calculation of a sequence could be processed concurrently, as each image's spatial descriptions are also independent. Once an image sequence's features are computed, adding another to the database is trivial. In this way it is possible to increase the database size over time, without the need for large compute power to re-analyse the database. This is a problem which hampered previous holistic shape description methods, [29].

7.1.3 Cartesian velocity moment selection

Once specific Zernike velocity moments have been isolated (through preliminary analysis or theoretical justification), a useful reduction in the complexity of the calculation can be utilised, by decomposing these selected Zernike velocity moments into their respective Cartesian components. Any future analysis (i.e. the addition of new subjects to a gait database), could be accomplished by expressing the Zernike velocity moments in terms of Cartesian ones (Section 3.4), providing a reduction in computational requirement and thus an increase in speed. This is effectively

using the Zernike velocity moments to aid the selection of less-correlated Cartesian velocity moments.

This approach could also be used to increase the understanding of the content of the two velocity moments. In general the traditional Cartesian moments are more widely understood (as compared to the Zernike moments), partly due to their simplicity, but also because they predate the Zernike moments. This knowledge could be used to help further understand the Zernike velocity moments. First the velocity moments (both Cartesian and Zernike) of an image sequence with known distribution are calculated. By then analysing both sets of results in parallel it may be possible to identify which moments characterise which specific features from the image sequence. This type of analysis would enable moment selection based on extracting specific information from the image sequence distribution. Equally this approach could be utilised to aid the analysis of the traditional Zernike moments.

7.1.4 Velocity moment content through reconstruction

Further analysis of the velocity moments may be possible by utilising moment reconstruction theory, specifically by using the velocity moments that proved successful for classification. Through moment reconstruction, an insight could be gained into which characteristics of the image sequence are captured by specific moments. However, in general, a greater number of moments will be needed for accurate image reconstruction than for classification. Thus, images reconstructed using a small number of selected moments will invariably not visually resemble the original image. There are two possible ways of implementing this reconstruction. The first would involve storing the corresponding velocity moment value for each image within the sequence (before summation), and then reconstructing a complete sequence of images, utilising the single image reconstruction theory - Sections 2.2.3 and 2.4.3. Using these reconstructed images, the actual gait characteristics extracted by the velocity moments could be studied. Figure 7.1a shows example images from an HiD ST sequence. Figure 7.1b shows the corresponding Zernike velocity moment reconstructed versions, using just those velocity moments used for classification (Table 5.24) and Figure 7.1c shows the thresholded reconstructions. The images are visibly different, while the reconstructed image corresponding to the subject's 'legs together' stance has a symmetrical nature.

The second method would use the theory in Sections 3.2.1 and 3.3.2 to reconstruct a single image, from the velocity moments calculated from the complete sequence. By producing a single composite image per sequence, it may be possible to again visibly observe the differences between subjects, which are being encoded by the velocity moments. Figure 7.2 shows three example Zernike velocity moment reconstructed images. Two are produced from subject sequences from HiD

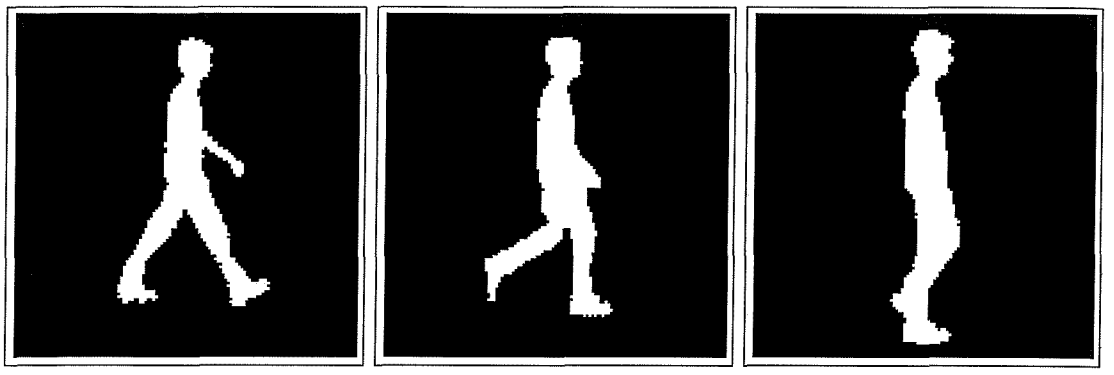
ST database, while the third is from a sequence in the UCSD ST database. Once again, visible differences between the three images are apparent. The moments used for this reconstruction were those selected for the HiD ST classification, detailed in Table 5.24. (The results shown in Figure 7.1 correspond to the same subject as the results in Figure 7.2a.) This type of reconstructed image could be used for classification by another method eg. a Fourier transform. Alternatively, this method of reconstruction could be used to perform the initial velocity moment selection for alternative applications, i.e. class separation within the database via extraction of a specific visual characteristic from each image sequence. Finally, these kinds of analyses may provide further information regarding the existence of symmetry (or asymmetry) within the human gait cycle.

7.1.5 Zernike mapping and optimal encoding

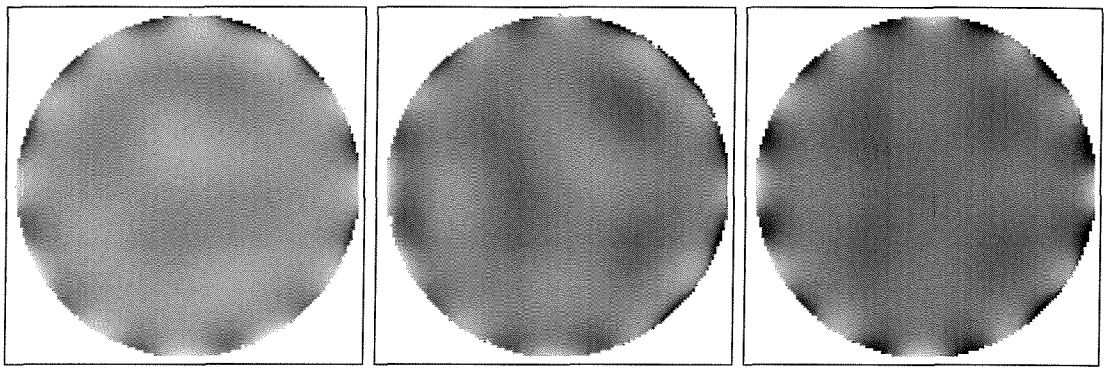
The Zernike (pre-processing) mapping function (Equation 2.58) is used to enable scale and translation invariant descriptions when mapping the shape onto the unit disc, where the scale of the mapped shape is set by β. As we have already seen the radial polynomials become more efficient at encoding image detail as r approaches unity (as detailed in Section 2.4.2 and shown in Figures 2.6 and 2.7). Considering the Zernike moments, descriptions closer to the unit disc's circumference will gain greater weighting than those about the origin. The pre-set value of β is very much dependent on the shape's distribution, essentially making it application specific. For this study its value was altered through manual observation to ensure that the shape (most often a subject's silhouette) did not interact with the perimeter of the unit disc, while maintaining a shape whose extremities were approaching 90% of the unit disc's radius ($r = 0.9$). Thus, determining a value for β enabling application-specific optimum encoding has not been considered, a possible area of further investigation for both the Zernike velocity moments and traditional Zernike moments.

7.1.6 Tailored basis functions

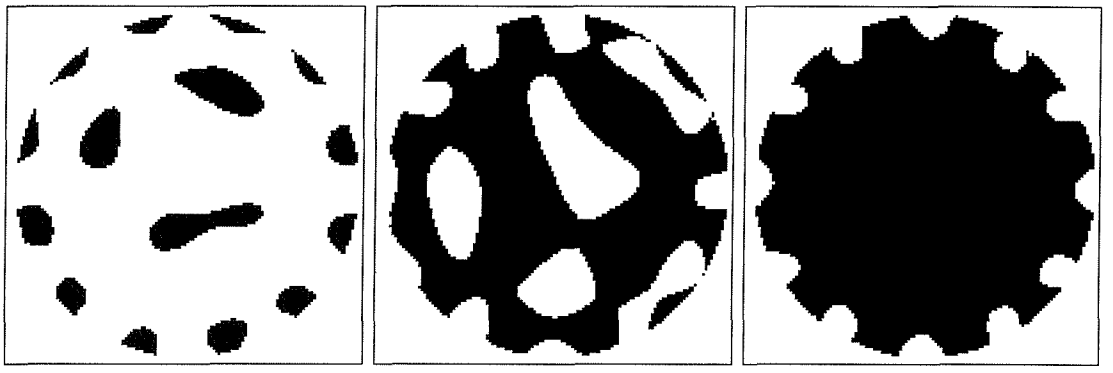
Currently the work has been concerned with the use of two basis functions, Cartesian and Zernike, producing the spatial descriptions within the velocity moments. The Zernike polynomials have the advantage of the orthogonality property producing less correlated descriptions. The Cartesian basis, has the advantage of computational simplicity, coupled with the disadvantage of the correlated non-orthogonal description. However, it is possible to design a specific basis function, tailored to the application. If we consider human gait, then it could prove prudent to use a basis function which specifically encodes sinusoidal motions (arms, legs, hip rotations



(a) Example images from a HiD ST sequence.



(b) Their reconstructed versions (using 8 velocity moments).



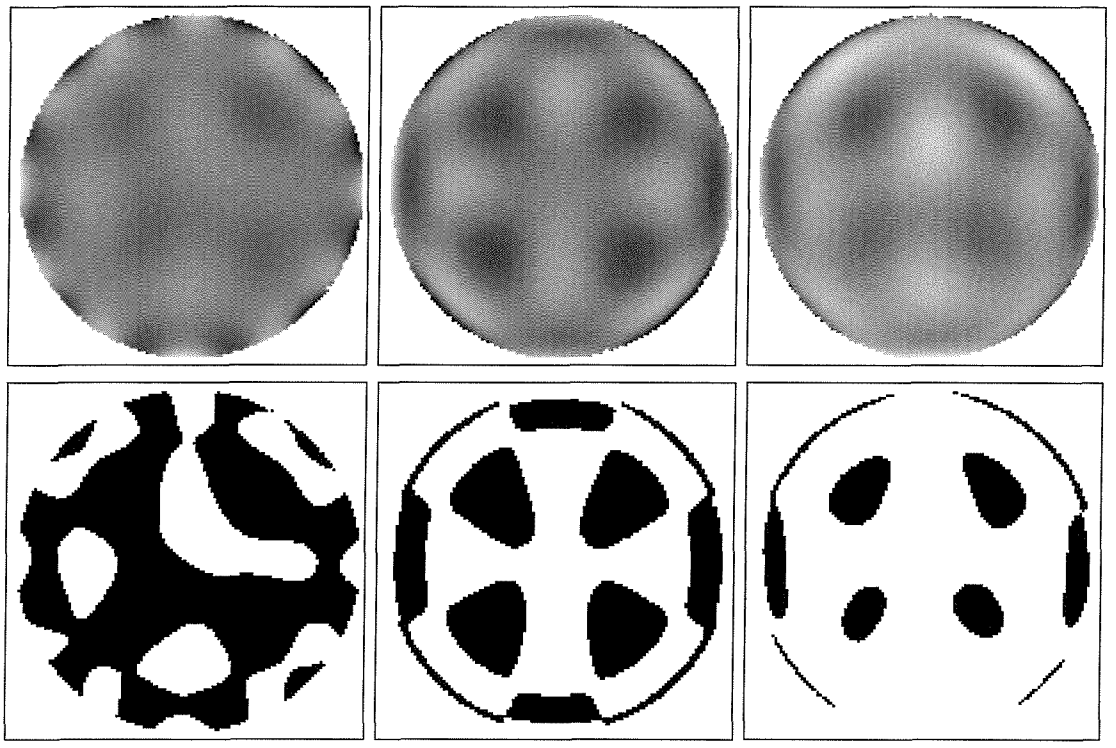
(c) Their thresholded reconstructed versions.

Figure 7.1: Example images from a HiD ST database sequence along with their corresponding reconstructed versions.

etc throughout the image sequence). For example, S in Equation 3.1 could be:

$$S(i,p,q) = \sin^p(x - \overline{x_i}) \sin^q(y - \overline{y_i}) \tag{7.1}$$

This will produce descriptions which are bounded in size to unity, and highly correlated due to the relationship between $\sin(x)$, $\sin^2(x)$, $\sin^3(x)$ etc which can be seen in Figure 7.3a. It may prove useful in detecting the presence of articulated (sinusoidal) motion, within a sequence (i.e. the difference between a sequence of a person walking, and a sequence of a moving car - see also Section 7.2.4). However, many more orthogonal basis functions exist including Hermite, Chebyshev, Laguerre and



(a) HiD Male. (b) HiD Female. (c) UCSD Male.

Figure 7.2: Example Zernike velocity moment (complete sequence) reconstructed images and thresholded versions for three different subjects.

Jacobi. Alternatively, a specific orthogonal basis set could be designed [85]. For example the function:

$$y_n(x) = \sin nx \quad ; \quad n = 1, 2....\infty \tag{7.2}$$

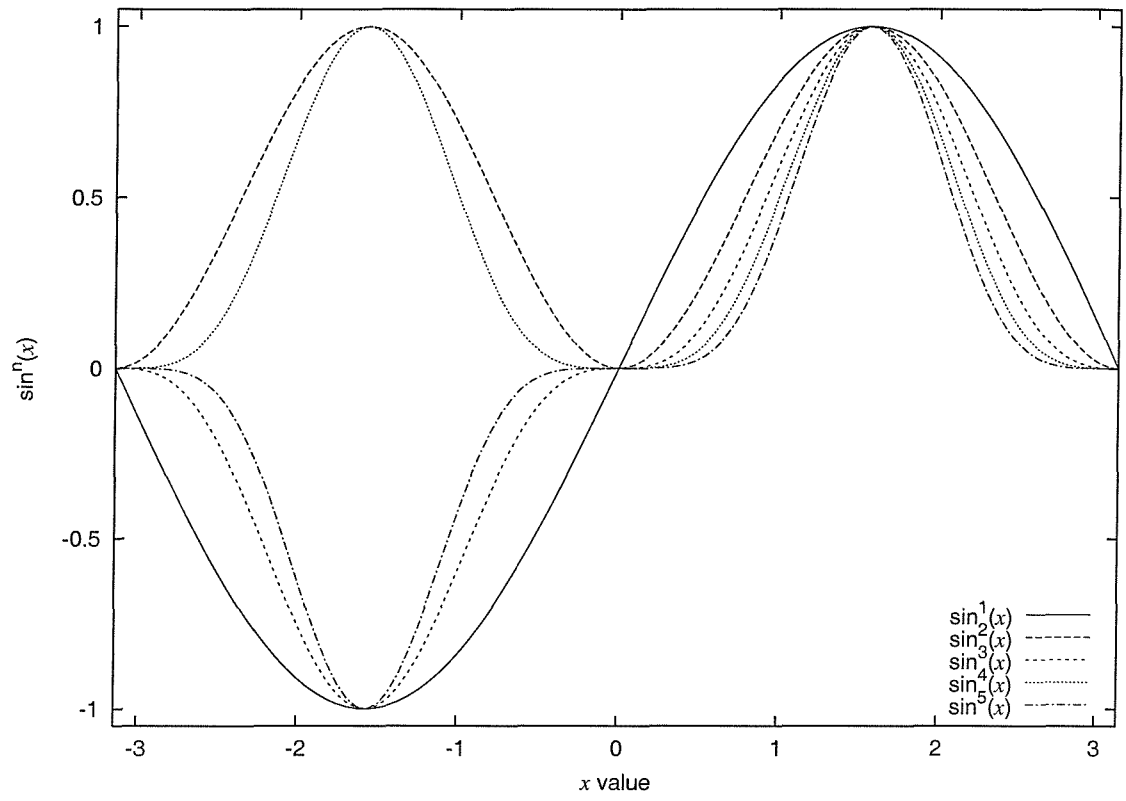
forms an orthogonal set on the interval $-\pi \leq x \leq \pi$ as shown in Figure 7.3b, suggesting that using:

$$S(i, p, q) = \sin\left(p(x - \overline{x_i})\right) \sin\left(q(y - \overline{y_i})\right) \tag{7.3}$$

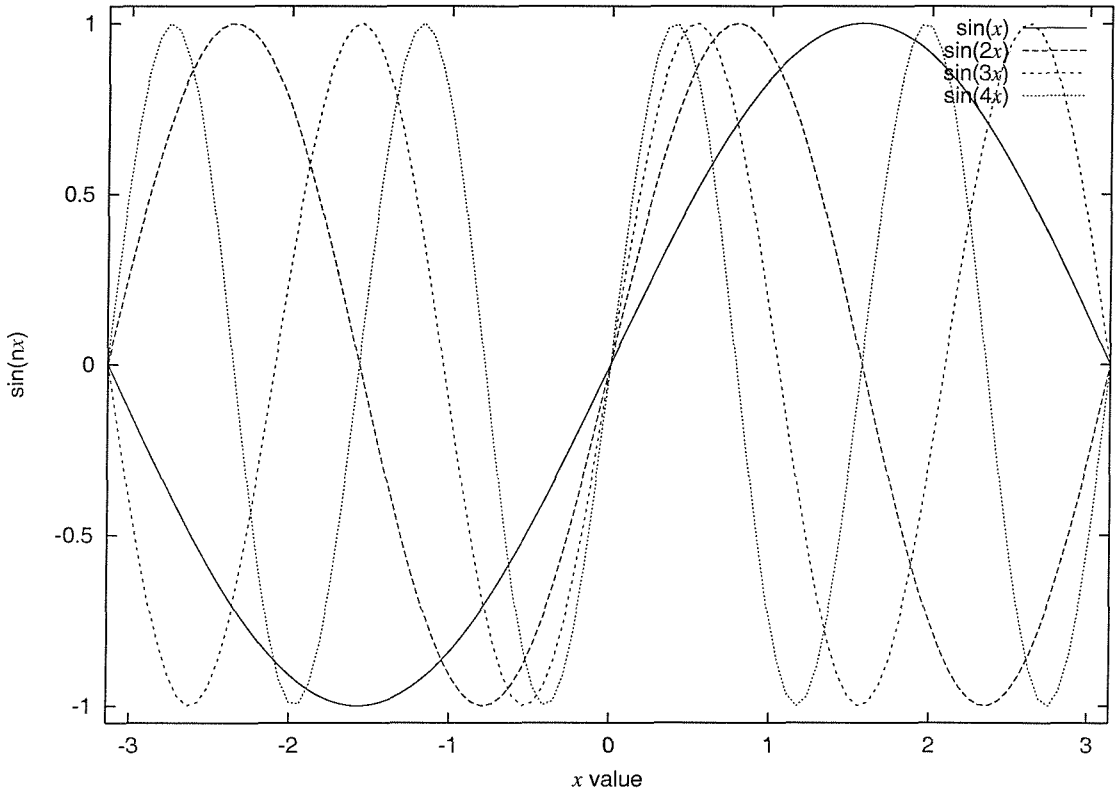
may prove productive, in terms of less correlated but bounded descriptors (equally cosines could be used in Equations 7.1 and 7.3 in place of the sine terms).

7.1.7 Cumulants

A normalised Gaussian distribution is completely characterised by its two moments, mean and standard deviation (variance). Non-Gaussian distributions require, in general, an infinite number of moments or cumulants to characterise them. Cumulants are a higher order statistic which can be easier to manipulate than moments as they scale linearly (L). (Moments scale with order n, $O \sim (L^n)$). This linearity makes cumulants more desirable when trying to model a distribution and is reflected



(a) $\sin^n(x)$



(b) $\sin(nx)$

Figure 7.3: Example alternative basis functions.

in their generating function, defined as the logarithm of the characteristic function (Equation 2.4):

$$G(w) = log[X(w)] = \sum_{n=1}^{\infty} \frac{\kappa_n(iw)}{n!} \tag{7.4}$$

expressed here as a Taylor series. The first three cumulants are equal to the first three centralised moments (or Cartesian moments of a zero mean distribution), while equations exist to convert between higher-order moments and cumulants [42]. It may prove fruitful to apply a similar framework including velocity to cumulants (as developed here for moments). Alternatively, the conversion equations between moments and cumulants may be applicable to the Cartesian velocity moments. Combining this with the work proposed in Section 7.1.3 may produce linearly scaled, orthogonal descriptions from the Cartesian velocity moments - through combinations of selected Cartesian velocity moments. Note that while the cumulants and moments are linked trivially for a single distribution, this is not the case when averaging over an ensemble of distributions.

7.1.8 Three-dimensional velocity moments

The reconstruction and classification of three dimensional objects is attracting increasing attention in recent times. Scene reconstruction through perspective volume intersection [40, 48, 60] (using multiple two-dimensional images), allows for three-dimensional scene analysis. The velocity moment framework is currently defined within a two-dimensional Cartesian coordinate system. However, Cartesian and centralised moments, and algebraic invariants have already been extended to three-dimensional space [69]. Extending the velocity moments to describe a three-dimensional space, shown here for the Cartesian basis would produce:

$$vm_{pq\mu\gamma\zeta} = \sum_{i=2}^{I} \sum_{x=1}^{M} \sum_{y=1}^{N} \sum_{z=1}^{O} U(i,\mu,\gamma,\zeta)\ S(i,p,q,r)\ P_{i_{xyz}} \tag{7.5}$$

where, here the sampled space is divided up into three-dimensional voxels $P_{i_{xyz}}$ (instead of two-dimensional pixels $P_{i_{xy}}$) and $S(i,p,q,r)$ arises from the centralised moments, in x, y and z:

$$S(i,p,q,r) = (x - \overline{x_i})^p (y - \overline{y_i})^q (z - \overline{z_i})^r \tag{7.6}$$

while $U(i,\mu,\gamma,\zeta)$ introduces three-dimensional velocity as:

$$U(i,\mu,\gamma,\zeta) = (\overline{x_i} - \overline{x_{i-1}})^\mu (\overline{y_i} - \overline{y_{i-1}})^\gamma (\overline{z_i} - \overline{z_{i-1}})^\zeta \tag{7.7}$$

Possible uses of this implementation include three-dimensional shape trajectory description and classification.

7.2 Application

7.2.1 Multiple shapes

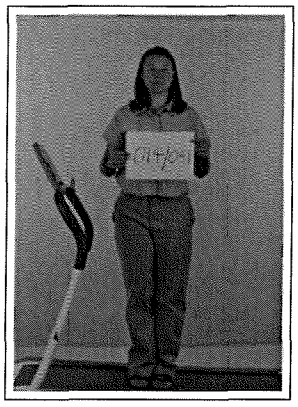
The work presented here has concentrated on the application of the velocity moments to images containing a single shape. The analysis could be extended to the description of multiple shapes within the field of view, describing the shapes, their motion and interaction with each other. (The problem of multiple shapes in a scene is also an extraction problem, as multiple shapes independently extracted are a set of single shapes.) Two identical rigid shapes moving away from each other at the same velocity will produce an average velocity of zero (as the COM will remain in the middle of the two shapes - the identical motion information of the same two shapes when stationary). Thus, describing the interaction between masses in a supernova (the explosion of most of the material in a star) may prove informative in terms of the distribution of energy, suggesting the approximate position of origin of the explosion. Alternatively, the analysis of a single deforming shape i.e. a bird, could be duplicated and combined together in such a way, as to help decompose and analyse a scene of flocking birds.

7.2.2 Human gait - Gender and age classification

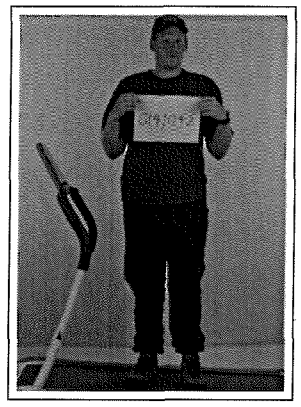
Murray showed that the amount of upper body sway was greater for males than females, [55]. When viewed from the side this movement is visible as a vertical motion or 'bobbing' as the subject walks, [13]. Velocity moments characterising y axis motion (or spread) may be able to detect this. Cutting [13] also found that gender identification by humans may depend on structural differences. The ratio of hip to shoulder width measurements for males and females was found to differ, a characteristic which may be detectable using the velocity moments from frontal viewing of the subject. Figure 7.4 demonstrates these differences, wider shoulders dominate the male stature, whereas the female has more slender shoulders and more pronounced hips. When shoulder width is divided by hip width, males produce a higher ratio. The shoulder to hip ratios for the subjects in Figure 7.4 are 0.96 (female) and 1.15 (male), consistent with Cutting's findings. Equally, recent work has suggested ways of distinguishing adults from children using stride length and frequency information [14]. It may be possible to use velocity moments describing spread in the x axis to further these ideas.

7.2.3 Animal movement analysis

One apparent extension from the analysis of human gait, is to apply the same techniques to animal gait. Preliminary analysis of animal gait classification includes work using symmetry [26] and mask operators [20], both producing promising results. Here the velocity moments could be used to discriminate between quadrupeds and bipeds, both within an animal dataset and between animal and human datasets.



(a) Ratio of 0.96.



(b) Ratio of 1.15.

Figure 7.4: A female (a) and male (b) subject, viewed from the front demonstrating variations in hip to shoulder ratios.

Further, the discrimination between a combination of the two may be possible, i.e. the difference between a horse and a horse with rider. These ideas may prove useful in video-sequence database browsing. Figure 7.5 shows exemplar animal silhouette data.

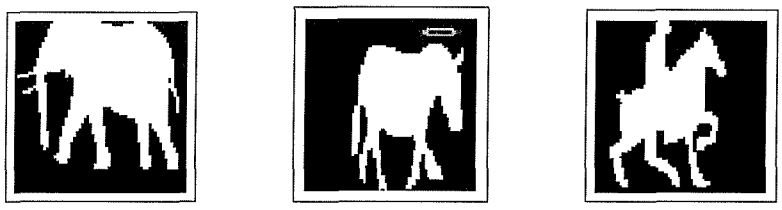


Figure 7.5: Example images from an animal database, an elephant, a zebra and a hoarse (with rider) respectively.

7.2.4 Types of motion - Trajectory description

If a (non-windowed) image sequence of a moving shape is processed by the velocity moments then the description will include the overall average velocity of that shape, along with detailed velocity information from consecutive images. In this way it may be possible to recognise a sequence of movements, such as a subject walking into a room, stopping and then exiting or a subject walking slowly and then running away. This may mean paying attention to structural velocity moments (eg. A_{0200}) to recognise individuals, while looking at specific velocity terms (eg. A_{**10} or A_{**20}) to study the type of motion (refer to Section 3.2.4), or a mixture of the two. This analysis of trajectories could be applied to gesture recognition or more explicitly - gestures through motion. A further aspect of this is the possibility of distinguishing between articulated and non-articulated motion, where here a possible concern would be to distinguish between road traffic (non-articulated) and pedestrians (articulated). An application for this work could be as an early warning system for the

car to slow down if a pedestrian steps into its path. This would involve identifying particular velocity moments (or characteristics) which are significantly different between the two types of motion. Areas of possible attention may concentrate on the differences, or lack of symmetry (or asymmetry) within the temporal sequences.

7.2.5 Alternative uses within computer vision

Here we are using the ANOVA technique to identify which moments are useful in terms of a shape classification problem. However, shape classification is only one area of image processing where moments are applicable. Other possible areas where the velocity moments may be applicable include: medical marker-less gait analysis, motion pose estimation (a variation of Section 7.2.4), image sequence encoding for transmission (utilizing reconstruction) or even motion template matching. These are all application areas which traditional moments are capable of analysing - if described by a single image. However, here it may be possible for the velocity moments to utilise the inclusion of motion.

Chapter 8

Conclusions

This final chapter draws together the results and conclusions of all the chapters which precede it. It begins with a brief summary of the work conducted throughout this study and finishes with overall conclusions.

8.1 Summary of work

Traditional statistical moment theory has concentrated on analysing single images. The motivation for this research has been to enable the analysis of temporal image sequences using statistical moments, with the drive towards producing a shape description which includes information about both shape and motion. It was proposed that by analysing a sequence of images, rigid and non-rigid shapes could be described in terms of their spatial characteristics and their motion, exploiting any temporal correlation present within the image sequence.

A simple framework called velocity moments was presented utilising the shape's centre of mass (COM) information to enable motion descriptions. The initial choice of spatial description function was the non-orthogonal Cartesian centralised moment's basis. This produced a simple temporal statistical moment technique. Via the use of Zernike polynomials the extension of the framework to enable orthogonal spatial descriptions was achieved. Thus, two techniques are presented, the non-orthogonal Cartesian velocity moments, and the orthogonal Zernike velocity moments. Further extensions and theory for these two techniques have been discussed including the topics of reconstruction, scale invariance issues, rotation invariant features, frame-rate and sequence length invariance and the conversion of features between the two techniques. To more fully demonstrate the properties of the velocity moments, they have been applied to the problem of human gait classification. Simple feature extraction techniques were utilised to produce features suitable for analysis by the velocity moments. A method based on human perception of gait has been proposed utilising both a subject's spatial silhouette and their movement, or optical flow. Seven gait databases were analysed producing

good classification results. The largest of these consisted of 50 subjects, a total of $\simeq$ 6000 images. Feature selection, or 'thinning' was achieved through the Analysis of Variance (ANOVA) tools, while the final classification used a simple k nearest neighbour approach. However, the main drive of this research was to produce features which enable classification of moving shapes, and not achieving optimal classification rates.

To further understand the performance characteristics of this new framework, performance analysis using the largest of the gait databases and the Zernike velocity moments was carried out. This analysis studied the issues of occlusion, image noise (simulated and 'real-world'), image resolution and time-lapse imagery, thus providing further understanding of the properties of the technique. Further possible extensions of the theory and techniques were then discussed, along with some preliminary results of these extensions. The extensions discussed include reconstruction, tailored basis functions, three-dimensional velocity moments and trajectory analysis. Finally, it is noted that a selection of the work detailed in this thesis exists in the following publications [71, 72, 73, 74].

8.2 Overall conclusions

A new description aimed at capturing both structural and temporal information of a time varying sequence has been proposed. It contains both scale and translation invariance. Two different variations of the technique have been presented, both yielding useful attributes. The Cartesian velocity moments are theoretically simplistic, although they will produce highly correlated features due to their non-orthogonal basis. This characteristic may hamper their successful use in the analysis of large datasets. The single-image orthogonality condition of the Zernike velocity moments means that the features produced are both smaller in magnitude than the Cartesian implementation and less correlated. They are however correlated in the sense that the images being described constitute a correlated sequence, a characteristic which has been shown to be beneficial. Further to this the Zernike velocity moments produce single-image scale invariant features, a property which is directly applicable to the problem of camera zoom on a piece of imagery.

It has been shown that the velocity moments have simple recognition properties, producing distinct results for different synthetic temporal test sets. The results reflected the structure of the velocity moment equation. They indicated that the expression holds information about the structure of the moving object as well as its velocity information. Hence, they can produce unique results for different objects moving at the same constant velocity, while, both unique *and* homogeneous results for the same object moving along different motion trajectories are possible. These

effects have been illustrated through the analysis of both rigid shapes (a bouncing ball) and non-rigid shapes (human gait case studies).

The Hu invariant moments have been seen to be inherently sensitive to perimeter noise, a result which is highly-dependent on the shape being described. This sensitivity is quite often due to the original (noise-free) value of the invariant moment being zero for the higher orders. Once noise is added the values for these moments begin to oscillate considerably, whereas the non-linear combinations of correlated Cartesian moments (comprising the Hu invariant set) appear capable of amplifying the effects of the perimeter noise. The method of velocity moments has been shown to have favourable characteristics when faced with the problems of image noise, simple occlusion, and loss in resolution. The performance of which is partly due to the integration of complete sequences, rather than describing each image separately. This suggests that the method would prove useful when applied to poorly extracted sequences, or possibly those where incomplete perimeter contours are apparent in images within a sequence. Increasing the length of the sequence will inherently enable improved description when hampered by these effects, a property not available for traditional moment analysis on single images. However, it has been explained how artificial performance analyses can themselves be misleading, especially in the case of image noise, while any area-based description method is essentially dependent on the performance of the initial feature extraction (pre-processing) technique. The importance of using significantly higher order moments for description has also been covered, even though higher order moments are more likely to be affected by noise.

Humans perceive gait by observing a person's overall shape and how this moves and changes as they walk. Using the velocity moments, a gait description method has been presented which takes these cues from nature. Classification by gait has been achieved using a person's build and stature. These results can be enhanced by including limb motion information (temporal templates containing optical flow), which describes a subject's intimate movements. The successful implementation of spatial template extraction techniques has in turn produced (visually) good temporal templates. These temporal templates should be more independent of the subject's surroundings and clothes than just extraction alone. Using these techniques high classification rates have been achieved on array of different human gait databases. However, possible specific velocity moments for gait recognition can only be determined through the analysis of larger databases, as the results presented here are by analysis of relatively small databases, in comparison with other biometrics. At the time of analysis the HiD database was the largest of its kind. However, this method of gait description has already produced promising results based on cues from human gait perception. Once an image sequence's features are computed,

adding another to the database is trivial. In this way it is possible to increase the database size over time, without the need for large compute power to re-analyse the database - a problem which hampered previous holistic shape description methods. A hypothesis concerning gait symmetry has been put forward, suggesting that the Cartesian velocity moments do indeed retain the symmetry properties exhibited by the centralised moments.

In conclusion, the velocity moments compress a temporal image sequence into a set of features that enable classification through both spatial and/or motion information. The use of an image sequence, in place of single images enables the exploitation of temporal correlation within the sequence, allowing the possibility of refining the description as the sequence length increases. The theory behind this new technique (and extensions to it) are presented, while its performance has been analysed using both synthetic data and through the application to human recognition by gait.

References

[1] J. K. Aggarwal and Q. Cai. Nonrigid motion analysis: Articulated and elastic motion. *Computer Vision and Understanding*, **70**(2):pp. 142–156, 1999.

[2] P. A. Beardsley, W. T. Freeman, D. B. Anderson, C. N. Dodge, M. Roth, C. D. Weissman, and W. S. Yerazunis. Computer vision for interactive computer graphics. *IEEE Computer Graphics I/O Devices*, **18**(3):pp. 42–53, 1998.

[3] S. O. Belkasim, M. Shridhar, and M. Ahmadi. Pattern recognition with moment invariants: A comparative study and new results. *Pattern Recognition*, **24**(12):pp. 1117–1138, 1991.

[4] C. BenAbdelkader, R. Cutler, H. Nanda, and L. Davis. EigenGait: Motion-Based Recognition of People Using Image Self-Similarity. *Proc. Audio- and Video-Based Biometric Person Authentication (AVBPA01)*, :pp. 284–294, 2001.

[5] C. BenAbdelkader, R. Cutler, H. Nanda, and L. Davis. Motion-Based Recognition of People in EigenGait Space. *Proc. Automatic face and gesture recognition (FGR02)*, :pp. 267–272, 2002.

[6] A. B. Bhatia and E. Wolf. On the circle polynomials of Zernike and related orthogonal sets. *Proc. Cambridge Philosophical Society*, **50**:pp. 40–48, 1954.

[7] H. Blum. A transformation for extracting new descriptors of shape. *Models for the perception of speech and visual form*, W. Wathem-Dum (Editor), Cambridge, Mass. MIT Press, 1967.

[8] H. Bulthoff, J. Little, and T. Poggio. A parallel algorithm for real-time computation of optical flow. *Letters To Nature*, **337**(9):pp. 549–553, 1989.

[9] G. M. Clarke and D. Cooke. *A Basic Course in Statistics*, chapter 22, pages 520–546. Arnold, 1998.

[10] P. R. Cohen. *Empirical Methods for Artificial Intelligence*, chapter 6-7, pages 185–287. MIT Press, 1995.

[11] R. T. Collins, R. Gross, and J. Shi. Silhouette-based human identification from body shape and gait. *Proc. Face and Gesture Recognition (FGR02)*, pages pp. 366–371, 2002.

[12] D. Cunado, M. S. Nixon, and J. N. Carter. Automatic gait recognition via model-based evidence gathering. *Proc. AutoID '99: IEEE Workshop on Identification Advanced Technologies*, :pp. 27–30, 1999.

[13] J. E. Cutting, D. R. Proffitt, and L. T. Kozlowski. A Biomechanical Invariant for Gait Perception. *Journal of Experimental Psychology*, **4**(3):pp. 357–372, 1978.

[14] J. W. Davis. Visual categorization of children and adult walking styles. *Proc. Audio- and Video-based Biometric Person Authentication (AVBPA01)*, :pp. 295–300, 2001.

[15] J. W. Davis and A. F. Bobick. The representation and recognition of action using temporal templates. *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR97)*, :pp. 928–934, 1997.

[16] M-P. Dubuisson and A.K. Jain. Contour extraction of moving objects in complex outdoor scenes. *International Journal of Computer Vision*, **14**(6):pp. 83–105, 1995.

[17] S. A. Dudani, K. J. Breeding, and R. B. McGhee. Aircraft identification by moment invariants. *IEEE Trans. on Computers*, **C-26**(1):pp. 39–46, 1977.

[18] S. T. Eke-Okoro, M. Gregoric, and L. E. Larsson. Alterations in gait resulting from deliberate changes of arm-swing amplitude and phase. *Clinical Biomechanics*, **12**(7/8):pp. 516–521, 1997.

[19] J. P. Foster, M. S. Nixon, and A. Prugel-Bennet. New area based metrics for automatic gait recognition. *Proc. British Machine Vision Conference (BMVC01)*, :pp. 233–242, 2001.

[20] J. P. Foster, M. S. Nixon, and A. Prugel-Bennet. New area based metrics for gait recognition. *Proc. Audio- and Video-based Biometric Person Authentication (AVBPA01)*, :pp. 312–317, 2001.

[21] H. Freeman. Boundary encoding and processing. *Picture Processing and Psychopictorics*, Lapkin and Rosenfeld (Editors), New York Academic Press:pp. 241–306, 1970.

[22] N. Friedman and S. Russell. Image segmentation in video sequences: A probabilistic approach. *Proc. Uncertainty in Artificial Intelligence (UAI97)*, :pp. 175–181, 1997.

[23] G. H. Granlund. Fourier preprocessing for hand print recognition. *IEEE Trans. on Computers*, **C-21**, 1972.

[24] M. G. Grant, M. S. Nixon, and P. H. Lewis. Extracting moving shapes by evidence gathering. *Pattern Recognition*, **35**:pp. 1099–1114, 2002.

[25] 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 (AVBPA01)*, :pp. 272–277, 2001.

[26] J. B. Hayfron-Acquah, M. S. Nixon, and J. N. Carter. Recognising human and animal movement by symmetry. *Proc. International Conference on Image Processing (ICIP01)*, :pp. 290–293, 2001.

[27] M-K. Hu. Pattern recognition by moment invariants. *Proc. IRE (Correspondence)*, **49**:pp. 1428, 1961.

[28] M-K. Hu. Visual pattern recognition by moment invariants. *IRE Trans. on Information Theory*, **IT-8**:pp. 179–187, 1962.

[29] P. S. Huang, C. J. Harris, and M. S. Nixon. Recognising humans by gait via parametric canonical space. *Artificial Intelligence in Engineering*, **13**:pp. 93–100, 1999.

[30] P. S. Huang, C. J. Harris, and M. S. Nixon. Human gait recognition in canonical space using temporal templates. *IEE Proc. Vision and Image Signal Processing*, **146**(2):pp. 93–100, April 1999.

[31] P. S. Huang, C. J. Harris, and M. S. Nixon. Comparing different template features for recognising people by their gait. *Proc. British Machine Vision Conference (BMVC98)*, **2**:pp. 639–648, Sept. 1998.

[32] S. Jabri. Detection and delineating humans in video images. *MSc Thesis, George Mason University, Virginia*, :, 2000.

[33] S. Jabri, Z. Duric, H. Wechsler, and A. Rosenfeld. Detection and location of people in video images using adaptive fusion of color and edge information. *Proc. International Conference on Pattern Recognition (ICPR00)*, **4**:pp. 627–630, 2000.

[34] R. Jain, R. Kasturi, and B. G. Schunck. *Machine Vision*. McGraw-Hill, 1995.

[35] X. Y. Jiang and H. Bunke. Simple and fast computation of moments. *Pattern Recognition*, **24**(8):pp. 801–806, 1991.

[36] A. Y. Johnson and A. F. Bobick. A multi-view method for gait recognition using static body parameters. *Proc. Audio-and Video-Based Biometric Person Authentication (AVBPA01)*, :pp. 301–311, 2001.

[37] A. Khotanzad and Y. H. Hongs. Invariant image recognition by Zernike moments. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **12**(5):pp. 489–497, 1990.

[38] A. Khotanzad and J-H. Lu. Classification of invariant image representations using a neural network. *IEEE Trans. on Acoustics, Speech and Signal Processing*, **38**:pp. 1028–1038, 1990.

[39] N. Kiryati and D. Maydan. Calculating geometric properties from Fourier representation. *Pattern Recognition*, **22**(5):pp. 469–475, 1989.

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

[41] L. Lee and W.E.L. Grimson. Gait analysis for recognition and classification. *Proc. Automatic face and gesture recognition (FGR02)*, :pp. 155–162, 2002.

[42] V. P. Leonov and A. N. Shiryaev. On a method of calculation of semi-invariants. *Theory of Probability and its Applications*, **4**(3):pp. 319–328, 1959.

[43] B. C. Li. Applications of moment invariants to neurocomputing for pattern recognition. *PhD Dissertation, The Pennsylvania State University*, 1990.

[44] B. C. Li. A new computation of geometric moments. *Pattern Recognition*, **26**(1):pp. 109–113, 1993.

[45] J. Little and J. Boyd. Describing motion for recognition. *Proc. International Symposium on Computer Vision*, pages pp. 235–240, Nov. 1995.

[46] J. J. Little and J. E. Boyd. Recognising people by their gait: the shape of motion. *Videre*, **1**(2):pp. 2–32, 1998.

[47] Y. Lui, K. L. Schmidt, J. F. Cohn, and R. L. Weaver. Facial asymmetry quantification for expression invariant human identification. *Proc. International conference on Automatic Face and Gesture Recognition (FG02)*, :to be published, 2002.

[48] W. N. Martin and J. K. Aggarwal. Volumetric descriptions of objects from multiple views. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **PAMI-5**(2):pp. 150–158, 1983.

[49] D. Meyer, J. Denzler, and H. Niemann. Model based extraction of articulated objects in image sequences for gait analysis. *Proc. IEEE International Conference on Image Processing (ICIP98)*, **3**:pp. 78–81, 1998.

[50] D. Meyer, J. Posl, and H. Niemann. Gait classification with HMM's for Trajectories of Body Parts Extracted by Mixture Densities. *Proc. British Machine Vision conference (BMVC98)*, **2**:pp. 459–468, 1998.

[51] F. Mokhtarian and A. K. Mackworth. Scale-based description and recognition of planar curves and two-dimensional shapes. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **8**(1):pp. 34–43, 1986.

[52] Moving Pictures Expert Group (MPEG). http://www.cselt.it/mpeg/.

[53] R. Mukundan and K. R. Ramakrishnan. *Moment functions in image analysis - Theory and applications.* World Scientific, 1998.

[54] H. Murase and R. Sakai. Moving object recognition in eigenspace representation: gait analysis and lip reading. *Pattern Recognition Letters*, **17**:pp. 155–162, 1996.

[55] M. P. Murray. Gait as a total pattern of movement. *American Journal of Physical Medicine*, **46**(1):pp. 290–332, 1967.

[56] J. M. Nash, J. N. Carter, and M. S. Nixon. Dynamic Feature Extraction via the Velocity Hough Transform. *Pattern Recognition Letters*, **18**:pp. 1035–1047, 1997.

[57] J. M. Nash, J. N. Carter, and M. S. Nixon. Extraction of moving articulated-objects by evidence gathering. *Proc. British Machine Vision Conference (BMVC98)*, **2**:pp. 609–618, 1998.

[58] M. S. Nixon and A. S. Aguado. *Feature extraction and image processing.* Butterworth Heinmann, 2002.

[59] S. A. Niyogi and E. H. Adelson. Analyzing and recognising walking figures in XYT. *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR94)*, :pp. 469–474, 1994.

[60] H. Noborio, S. Fukuda, and S. Arimoto. Construction of the octree approximating three-dimensional objects by using multiple views. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **10**(6):pp. 769–782, 1988.

[61] A. Papoulis. *Probability, random variables, and stochastic processes*, chapter 5, pages pp. 86–124. McGraw-Hill, 1992.

[62] M. Pawlak. On the reconstruction aspects of moment descriptors. *IEEE Trans. on Information Theory*, **38**(6):pp. 1698–1708, 1992.

[63] N. Peterfreund. Robust tracking of position and velocity with kalman snakes. *IEEE Trans. Pattern Analysis and Machine Intelligence*, **21**(6):pp. 564–569, 1999.

[64] 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. Automatic Face and Gesture Recognition (FGR02)*, :pp. 137–142, 2002.

[65] R. J. Prokop and A. P. Reeves. A survey of moment-based techniques for unoccluded object representation and recognition. *CVGIP Graphical models and Image Processing*, **54**(5):pp. 438–460, 1992.

[66] Y. Ricquebourg and P. Bouthemy. Tracking of articulated structures exploiting spatio-temporal image slices. *Proc. International Conference on Image Processing*, 3:pp. 480–483, 1997.

[67] R. J. Roddis. Extending the snake model to incorporate velocity. *Mphil Thesis, University of Southampton, UK*, 2002.

[68] R. Rosales. Recognition of human actions using moment-based features. *Boston University Computer Science Technical Report*, **BU 98-020**, 1998.

[69] F. A. Sadjadi and E. L. Hall. Three-dimensional moment invariants. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **PAMI-2**(2):pp. 127–136, 1980.

[70] D. Shen and H. S. Ip. Discriminative wavelet shape descriptors for recognition of 2-d patterns. *Pattern Recognition*, **32**(2):pp. 151–165, 1998.

[71] J. D. Shutler and M. S. Nixon. Zernike velocity moments for the description and recognition of moving shapes. *Proc. British Machine Vision Conference (BMVC01)*, **2**:pp. 705–714, 2001.

[72] J. D. Shutler, M. S. Nixon, and C. J. Harris. Statistical gait recognition via velocity moments. *IEE Colloquium - Visual Biometrics*, (Digest 1999:):pp. 11/1–11/5, 1999.

[73] J. D. Shutler, M. S. Nixon, and C. J. Harris. Global statistical description of temporal features. *Proc. International Society for Photogrammetry and Remote Sensing Congress (ISPRS00)*, :pp. 720–726, 2000.

[74] J. D. Shutler, M. S. Nixon, and C. J. Harris. Statistical gait recognition via temporal moments. *Proc. IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI00)*, :pp. 291–295, 2000.

[75] N. K. Sinha. *Linear Systems*, chapter 3. Wiley, 1991.

[76] A. Sluzek. Identification and inspection of 2-d objects using new moment-based shape descriptors. *Pattern Recognition Letters*, **16**:pp. 687–697, 1995.

[77] N. J. C. Strachan, P. Nesvadba, and A. R. Allen. A method for working out the moments of a polygon using an integration technique. *Pattern Recognition Letters*, **11**:pp. 351–354, 1990.

[78] R. Takamatsu. A pointing device gazing at hand based on local moments. *SPIE*, **3028**:pp. 155–163, 1997.

[79] M. R. Teague. Image analysis via the general theory of moments. *Journal of the Optical Society of America*, **70**(8):pp. 920–930, 1979.

[80] C. Teh and R. T. Chin. On image analysis by the method of moments. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **10**(4):pp. 496–513, 1988.

[81] D. Thompson. *The Oxford Dictionary of Current English: Second edition.* Oxford University Press, 1992.

[82] J. M. Wilf and R. T. Cunningham. Computing region moments from boundary representations. *Jet Propulsion Laboratory publication, California Institute of Technology, Pasadena, CA*, 1979.

[83] J. Wood. Invariant pattern recognition: A review. *Pattern Recognition*, **29**(1):pp. 1–17, 1996.

[84] C. Y. Yam, M. S. Nixon, and J. N. Carter. Gait recognition by walking and running: A model-based approach. *Proc. Asian Conference on Computer Vision (ACCV02)*, **1**:pp. 1–6, 2002.

[85] L. L. Yudell. *Mathematical functions and approximations*, chapter 11, page 432. Academics Press Inc, 1975.

[86] C. T. Zahn and R. Z. Roskies. Fourier descriptors for plane closed curves. *IEEE Trans. on Computers*, **C-21**:pp. 269–281, 1972.

[87] F. Zernike. Beugungstheorie des Schneidenverfahrens und seiner verbesserten Form, der Phasenkontrastmethode (Diffraction theory of the cut procedure and its improved form, the phase contrast method). *Physica*, **1**:pp. 689–704, 1934.

Appendix A

Noise analysis

A.1 Perimeter noise

This appendix details the perimeter noise algorithm, as used in Section 3.2.3 and helps verify the results gained through experimentation via theoretical analysis of the noise function. Perimeter noise is applied to a sequence of binary images of a simple moving shape. This is achieved by first determining the perimeter of the shape in each image, and then applying zero-mean Gaussian noise to only those pixels located on the perimeter. The algorithm effectively moves the perimeter pixel as the noise affects a pixel's position, not its value.

A.1.1 Perimeter noise algorithm

Canny edge detection is first applied to each binary image, producing edge magnitude $E(x, y)$ and phase information $\phi(x, y)$. The phase information $\phi(x, y)$ is used to determine the direction in which the perimeter pixel moves. Using the edge (perimeter) image $E(x, y)$ as a mask - for every pixel in the original image which exists on the perimeter of the shape, a zero-mean Gaussian random number w is generated. Positive values of w cause a line to be drawn (of w pixels in length) along the orientation of the edge in a positive direction expanding the shape's perimeter. Negative values of w cause a line to be removed of length w, starting at the perimeter pixel location, again in the orientation of the edge, causing the perimeter to shrink into the shape. The amount of noise applied to each sequence can be adjusted by altering the variance σ^2 of the Gaussian distribution. Figure A.1 shows an example image from the tug (boat) sequence of images, along with the corresponding perimeter mask and perimeter noise images for $\sigma^2 = 2, 4$ and 6. It is noted that noise on a perimeter produced through poor extraction will tend to be more structured than just the single pixel-wide lines used here, i.e. in practice blocks of pixels along the perimeter are more likely to appear or disappear. However, we are interested in the average affect of altering the perimeter, allowing the use of this simplified model. Further, the detection of the perimeter could be achieved

(a) Original (b) Perimeter (c) $\sigma^2 = 2$ (d) $\sigma^2 = 4$ (e) $\sigma^2 = 6$

Figure A.1: Original shape, perimeter mask and example perimeter noise images.

by using a Sobel operator. This results in a thicker perimeter mask, allowing for parts of the perimeter to become detached from the shape (as the noise is applied to all pixels within the perimeter). Here we are only interested in deforming the perimeter (i.e. applying the noise to the perimeter pixel locations), hence the use of the Canny method. Although is it still possible for parts of the perimeter to become detached when using the Canny method. This occurs where large variance noise overlaps, occurring at areas of high frequency in the perimeter mask (rapid changes in the shape's perimeter orientation).

A.1.2 Perimeter noise analysis of the Cartesian velocity moments

The central limit theorem [61] states that given a population distribution, the distribution of samples about it's mean approaches a normal, or Gaussian distribution, given enough samples. The larger the number of samples the better this approximation becomes. In this limit we can assume all noise (perimeter or otherwise) to be Gaussian distributed. However, there will always be outliers to any distribution, while in practice here we are actually using a discrete approximation to a Gaussian. We are attempting to simulate zero-mean Gaussian noise around the perimeter of the shape. Due to the zero-mean condition it is possible to both add and remove pixels. By considering the effect of the perimeter noise on the centralised moments, it is then possible to determine the effect of the perimeter noise on the velocity moments. Referring first to the centralised moments, defined by Equation 2.19, where the COM for the x co-ordinate is:

$$\overline{x} = \frac{m_{10}}{m_{00}} \tag{A.1}$$

substituting in for m_{10} and m_{00} using Equation 2.19 gives:

$$\overline{x} = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N} xP_{xy}}{\sum_{x=1}^{M}\sum_{y=1}^{N} P_{xy}} \tag{A.2}$$

Considering the noise on the x co-ordinates of the perimeter expressed as $(x+\delta)$ produces:

$$\begin{aligned}\overline{x} &= \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}(x+\delta)P_{xy}}{\sum_{x=1}^{M}\sum_{y=1}^{N}P_{xy}} \\ &= \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}xP_{xy}+\delta P_{xy}}{\sum_{x=1}^{M}\sum_{y=1}^{N}P_{xy}}\end{aligned} \tag{A.3}$$

On average the image mass ($\sum_{x=1}^{M}\sum_{y=1}^{N}P_{xy}$) will be unchanged by a zero mean random process. Since the image and noise are uncorrelated then:

$$E[\delta P_{xy}] = E[\delta]E[P_{xy}] \tag{A.4}$$

where $E[.]$ is the expectation. Since the Gaussian distributed noise is zero-mean, as the number of samples (of the discrete approximation) increases then:

$$E[\delta] \to 0 \tag{A.5}$$

Using this result and Equation A.4 reduces Equation A.3 to:

$$\overline{x} = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}xP_{xy}}{\sum_{x=1}^{M}\sum_{y=1}^{N}P_{xy}} \tag{A.6}$$

A similar result is produced for the COM of the y co-ordinate. Accordingly the noise does not affect the center positions, or COM calculations. It must be noted that the image itself has changed, whereas the mean ($\overline{x}$) has remained constant (minus any inaccuracies due to the discrete implementation). Using the result that the COM coordinates are unchanged by the perimeter noise, we can move onto the velocity moments themselves. First we consider the velocity moment vm_{0010} (referring to Equation 3.2 in Chapter 3):

$$vm_{0010} = \sum_{i=2}^{I}\sum_{x=1}^{M}\sum_{y=1}^{N}(\overline{x_i} - \overline{x_{i-1}})\,P_{i_{xy}} \tag{A.7}$$

We have already seen that the COM calculations are unaffected by the perimeter noise. This motion estimate vm_{0010} is the difference between consecutive COMs, therefore it is not unreasonable to conclude that this result will also be unaffected by the perimeter noise. As such neither vm_{0010} or vm_{0001} will vary when zero-mean Gaussian perimeter noise perturbs the relevant co-ordinate (x or y respectively). Experimental results reflect this conclusion, as shown in Figures A.2 and A.12 where a peak-to-peak variation of $< 8\,\%$ is apparent. The moment values have

been plotted in terms of the percentage deviation from the original no-noise value. Any discrepancies where the value can be seen to fluctuate slightly appear to be due to rounding errors, and the distribution of the noise function being a discrete approximation to a Gaussian. Understandably these affects will be greater for smaller images and motions. However, the motion element of the moment descriptor has not been affected by the perimeter noise. Using this result we can now consider the velocity moment vm_{2010}. From Equation 3.2 this is expressed as:

$$vm_{2010} = \sum_{i=2}^{I}\sum_{x=1}^{M}\sum_{y=1}^{N}(x - \overline{x_i})^2 \ (\overline{x_i} - \overline{x_{i-1}}) \ P_{i_{xy}} \tag{A.8}$$

The motion information is unaffected by the perimeter noise, assuming this we can consider just the spatial description part of Equation A.8 i.e. vm_{2000}. Applying the noise to this expression (on the pixel coordinates) as before (and using the result in Equation A.5) produces:

$$\begin{aligned} E[vm_{2000} + \delta] &= E[vm_{2000}] + E[(\delta^2 + 2x\delta - 2\overline{x_i}\delta)P_{i_{xy}}] \\ &= E[vm_{2000}] + E[\delta^2 P_{i_{xy}}] \end{aligned} \tag{A.9}$$

Now using this result we can return to analysing vm_{2010} producing, in terms of the perimeter noise:

$$E[vm_{2010} + \delta] = E[vm_{2010}] + E[\delta^2 P_{i_{xy}}] \tag{A.10}$$

Therefore the velocity moment's spatial descriptions will be affected by the perimeter noise. However, for a rigid shape these effects can be reduced by exploiting temporal correlation (explained in Section 1.2).

A.1.3 Perimeter noise analysis of the Hu invariant moments

For completeness a similar analysis was applied to the (time-averaged) Hu invariant moments. The Hu invariant moment I_2 (from Section 2.3) is:

$$I_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \tag{A.11}$$

As before we assume that due to the zero-mean Gaussian process, the image mass will be unchanged. While the shapes in the analysed image sequences (tug-boat and overlaid-shapes) have constant size. Therefore, the scale normalisation (Equation 2.20) will have no effect on this analysis and η_{pq} can be considered to be μ_{pq} (an un-normalised centralised moment). By using the centralised moment definition (Equation 2.19) and adding the noise in the same way as for the velocity moment case, the noise contributions of the first term, η_{20} can be calculated. Assuming the

same conditions as before and using Equations A.5 and A.4 produces:

$$E[\eta_{20} + \delta] = E[\eta_{20}] + \ E[\delta^2 P_{i_{xy}}] \tag{A.12}$$

similarly the second term in Equation A.11 (η_{02}) produces the same result. Intuitively these two results when placed into Equation A.11 will cancel each other out, however, we will retain them. Again, using the same conditions to evaluate the last term in Equation A.11 (η_{11}) results in the expression:

$$E[\eta_{11} + \delta] = E[\eta_{11}] + \ E[\delta^2 P_{i_{xy}}] \tag{A.13}$$

Once this expression is then placed into Equation A.11, it can be seen that the error due to the noise ($E[\delta^2 P_{xy}]$) will always be positive, due to the squared term. Collecting all the noise terms together we now have:

$$E[I_2 + \delta] = E[I_2] + (E[\delta^2 P_{i_{xy}}] - E[\delta^2 P_{i_{xy}}])^2 \ + \ 4E[\delta^2 P_{i_{xy}}]^2 \tag{A.14}$$

This will cause the value of I_2 to increase beyond its true value as the amount of perimeter noise increases. This amplification of the noise is mainly due to the squared term. This conclusion is reflected in the results gained through testing, shown in Figures A.8 and A.18. The results shown in Figures A.2 to A.11 are for the moving overlaid-shapes sequence, and Figures A.12 to A.21 show the results for the tugboat sequence. As before, the moment values have been plotted in terms of their percentage deviation from the original no-noise value. The amplified effects of the noise are a result of the non-linear combinations of centralised moments which comprise this invariant set. Again, it is noted that through the exploitation of temporal correlation in the image sequence, these effects can be reduced. It is worth noting that I_1 is less affected by the perimeter noise than the other Hu invariant moments, as shown in Figures A.7 and A.17. This is a direct result of I_1 not comprising of non-linear combinations of centralised moments (refer to Equation 2.37). The slightly increasing trends shown by these plots reflect both the discrete approximation of the Gaussian process and the effect of the perimeter noise spreading the shape's edge pixels. This analysis has shown that the Hu invariant moments are more likely to be affected by perimeter noise, in comparison to the Cartesian velocity moments.

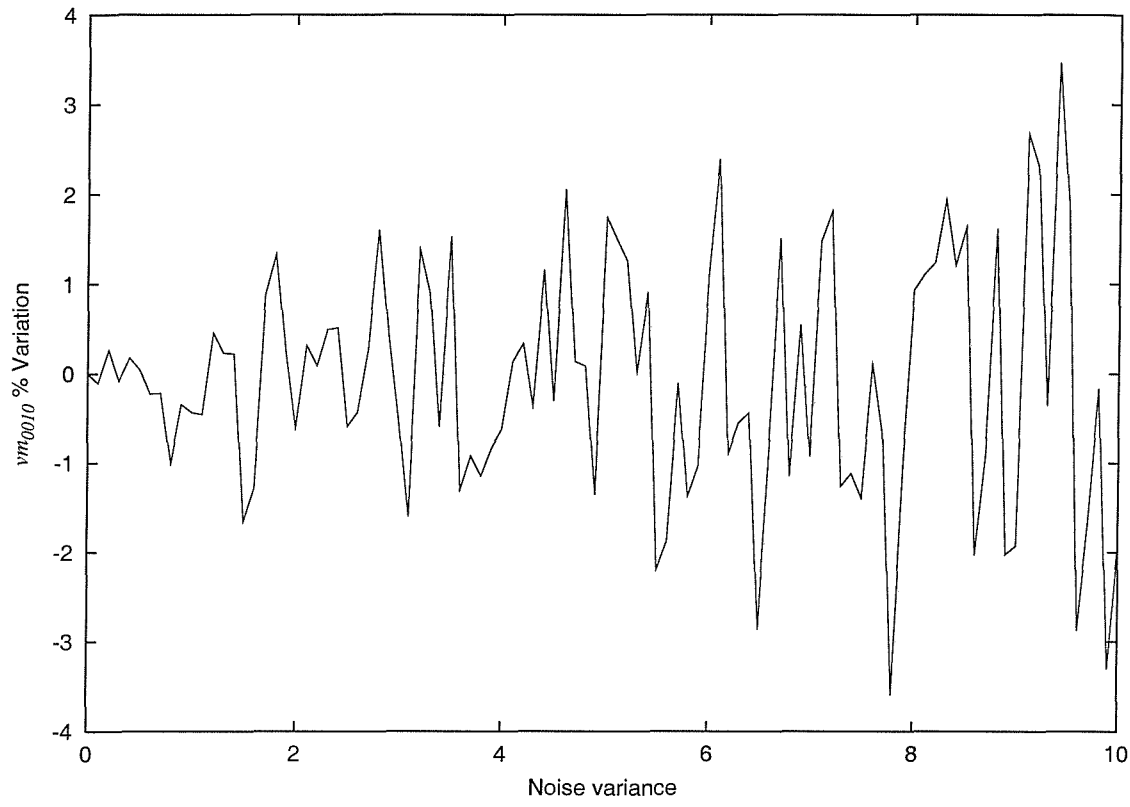


Figure A.2: Perimeter noise applied to the overlaid-shapes sequence - vm_{0010}

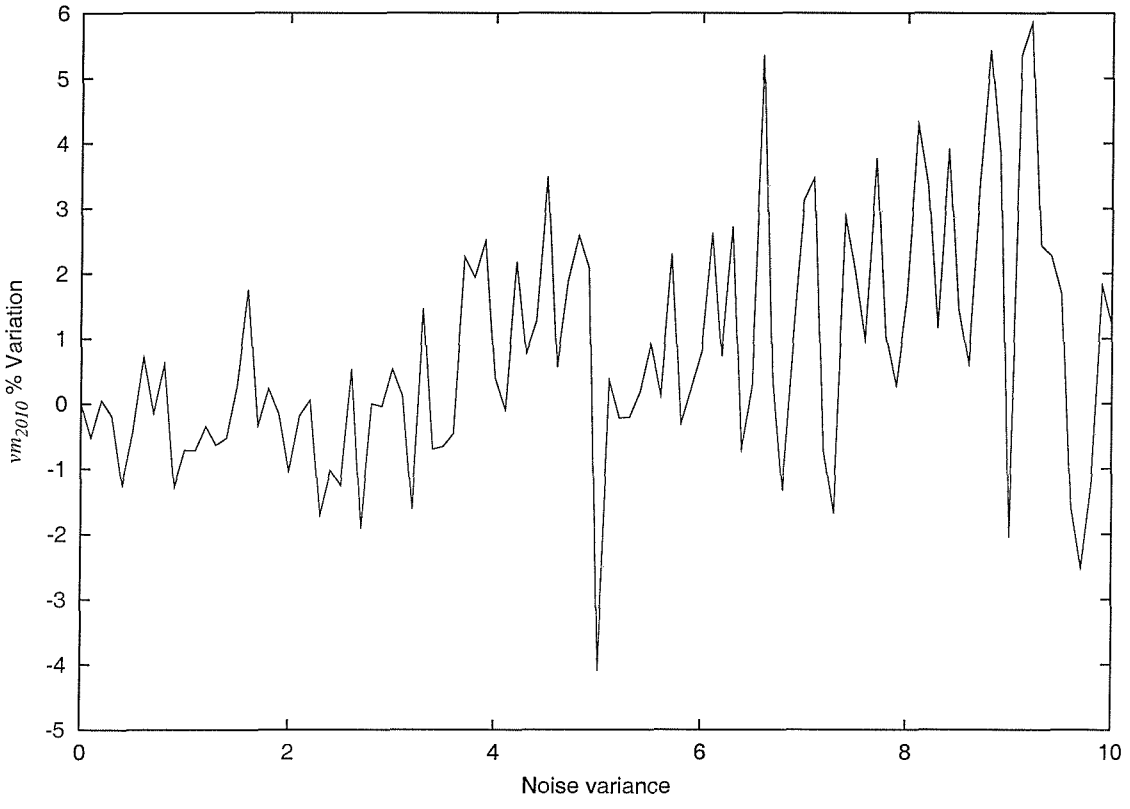


Figure A.3: Perimeter noise applied to the overlaid-shapes sequence - vm_{2010}

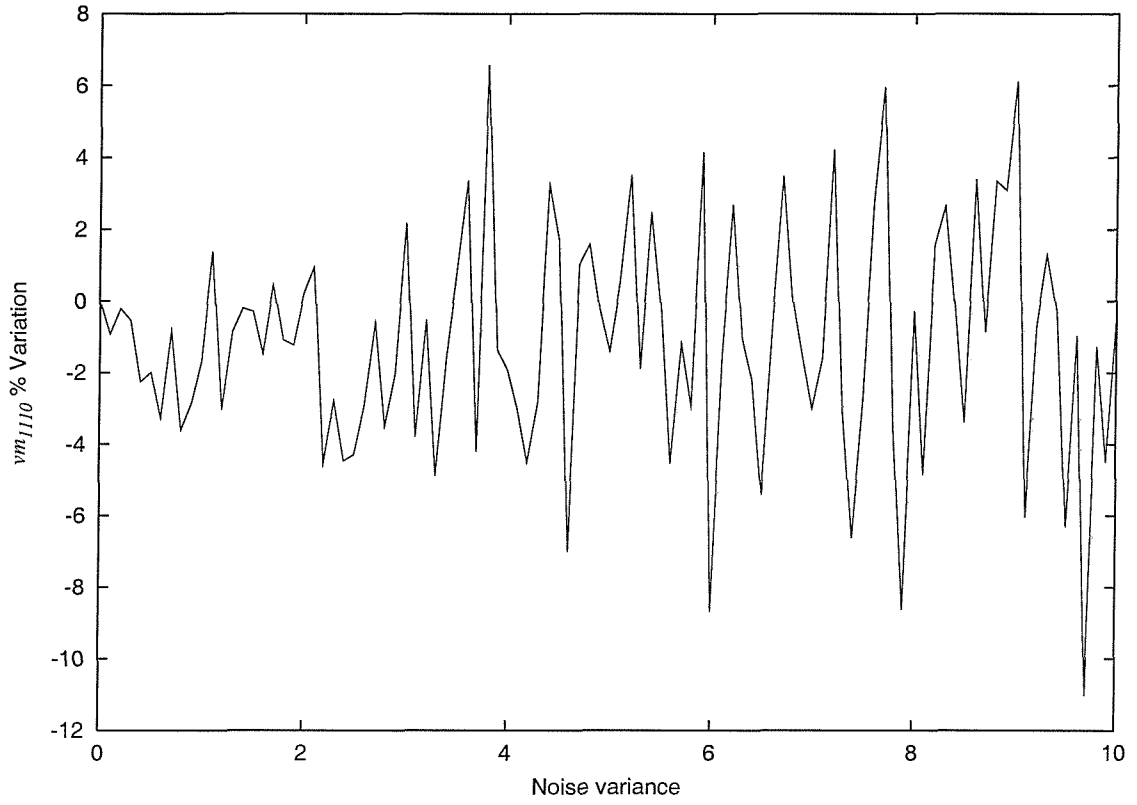


Figure A.4: Perimeter noise applied to the overlaid-shapes sequence - vm_{2210}

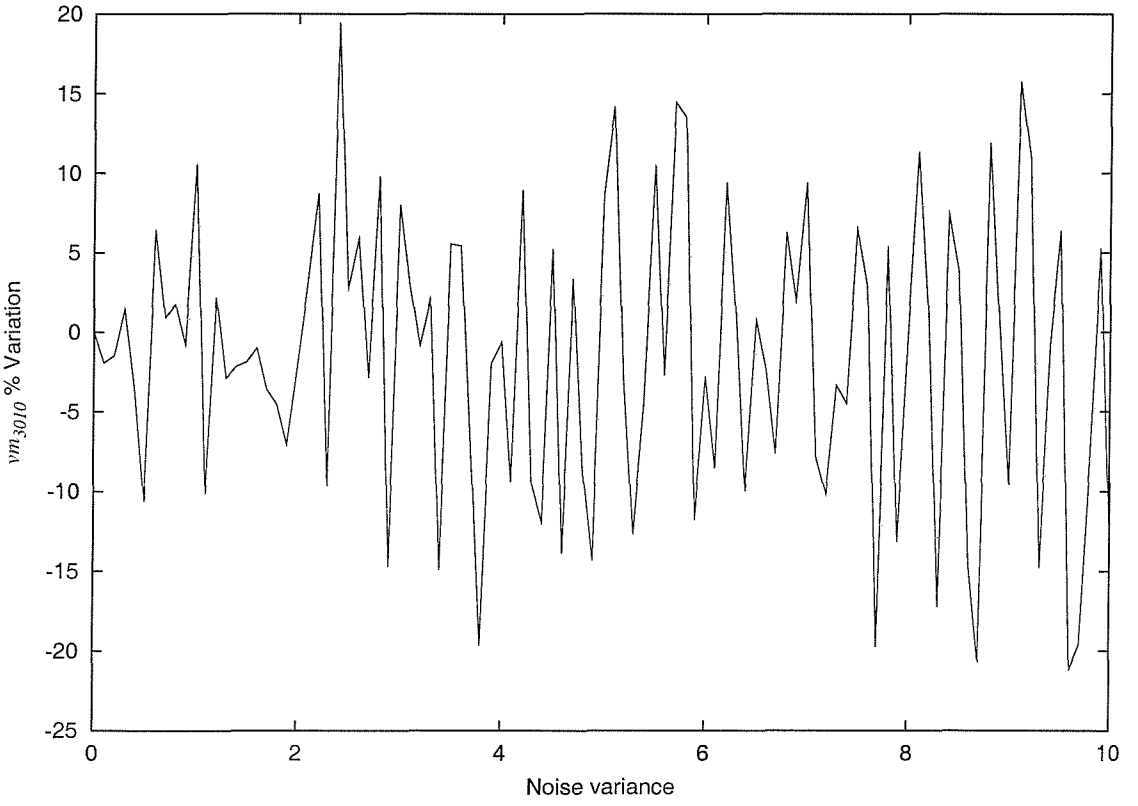


Figure A.5: Perimeter noise applied to the overlaid-shapes sequence - vm_{1110}

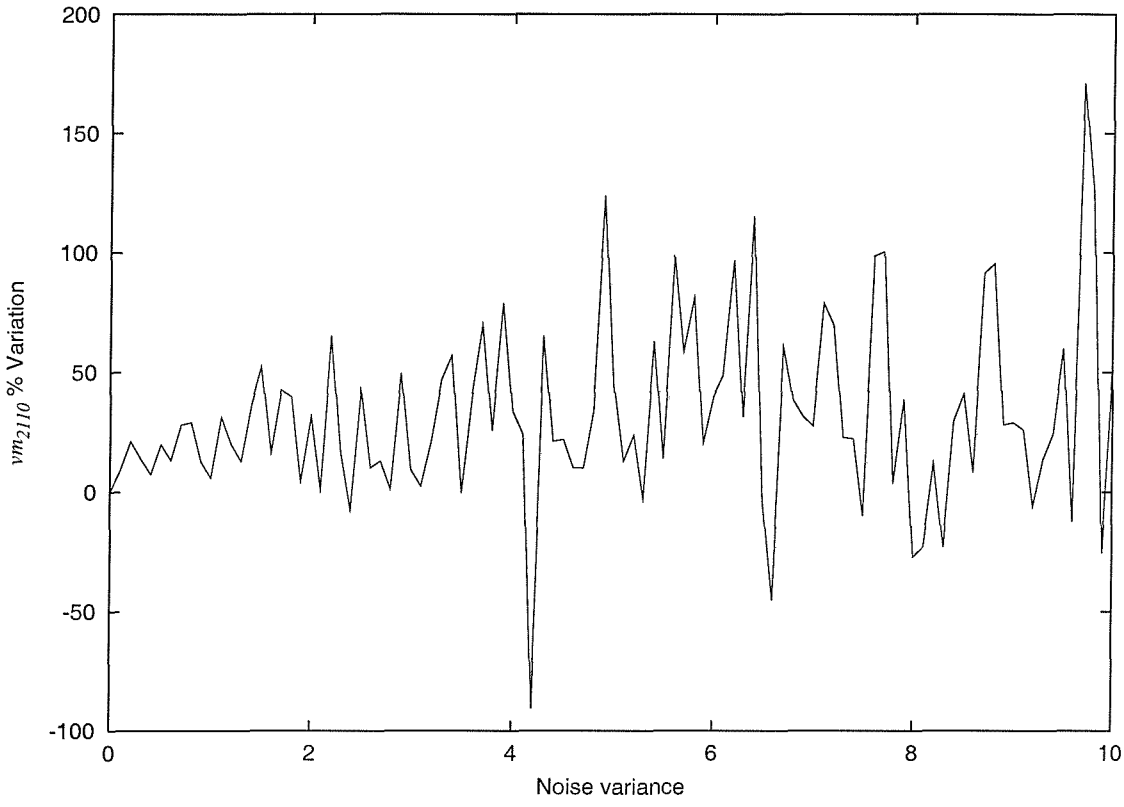


Figure A.6: Perimeter noise applied to the overlaid-shapes sequence - vm_{2110}

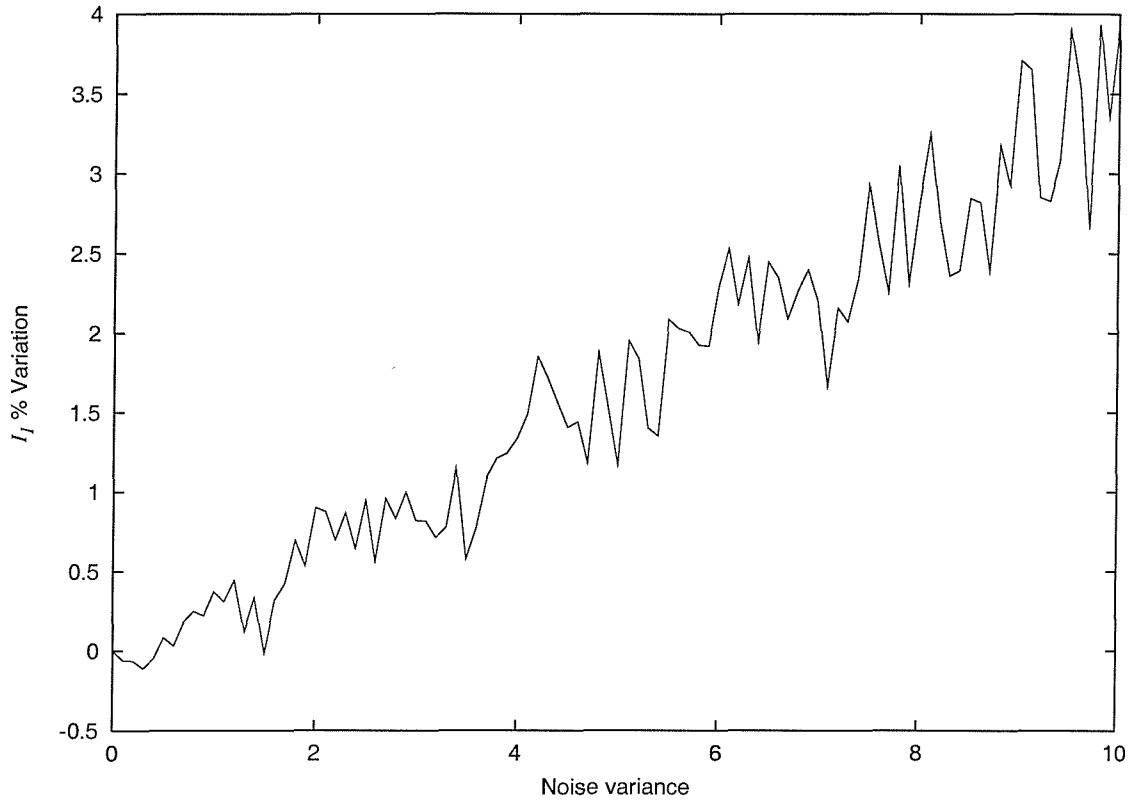


Figure A.7: Perimeter noise applied to the overlaid-shapes sequence - I_1

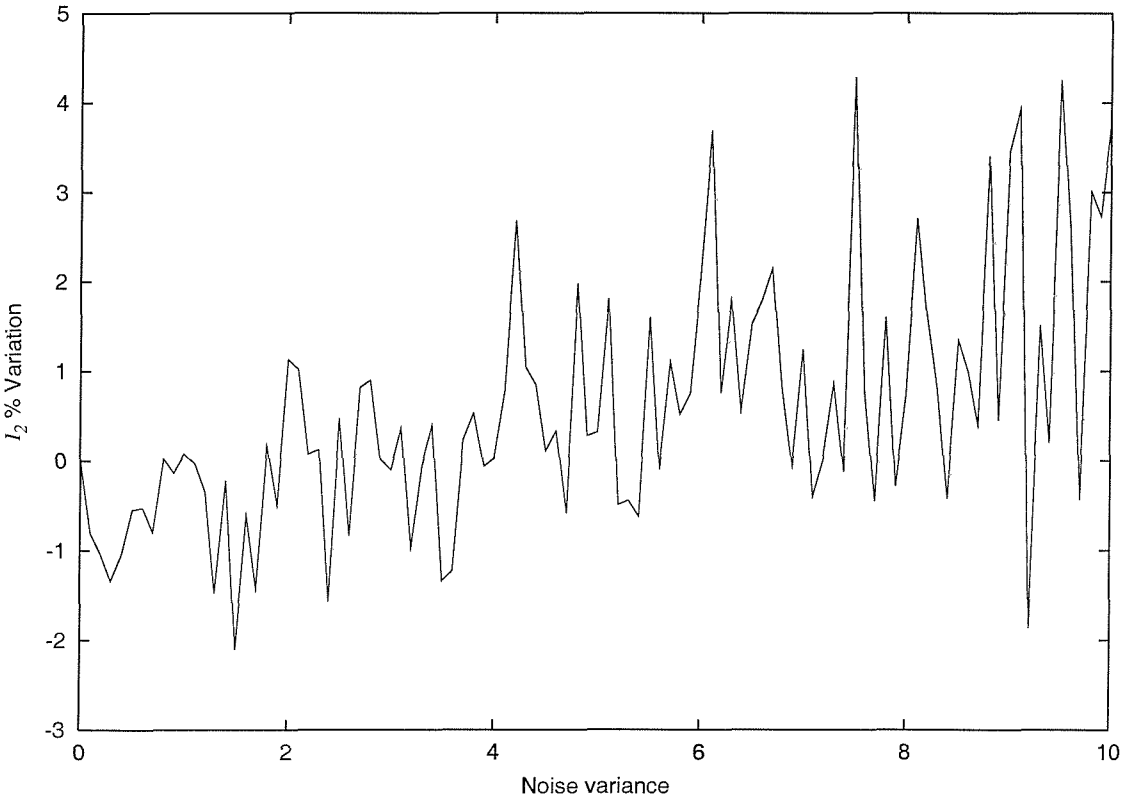


Figure A.8: Perimeter noise applied to the overlaid-shapes sequence - I_2

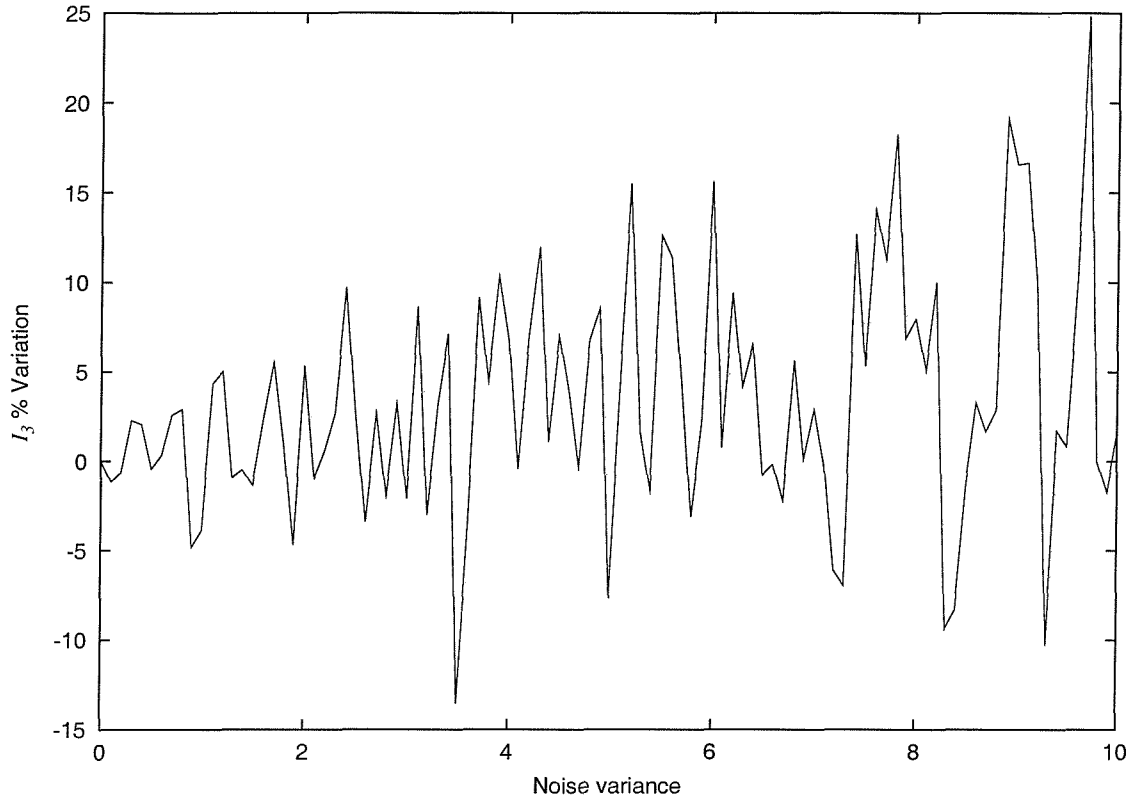


Figure A.9: Perimeter noise applied to the overlaid-shapes sequence - I_3

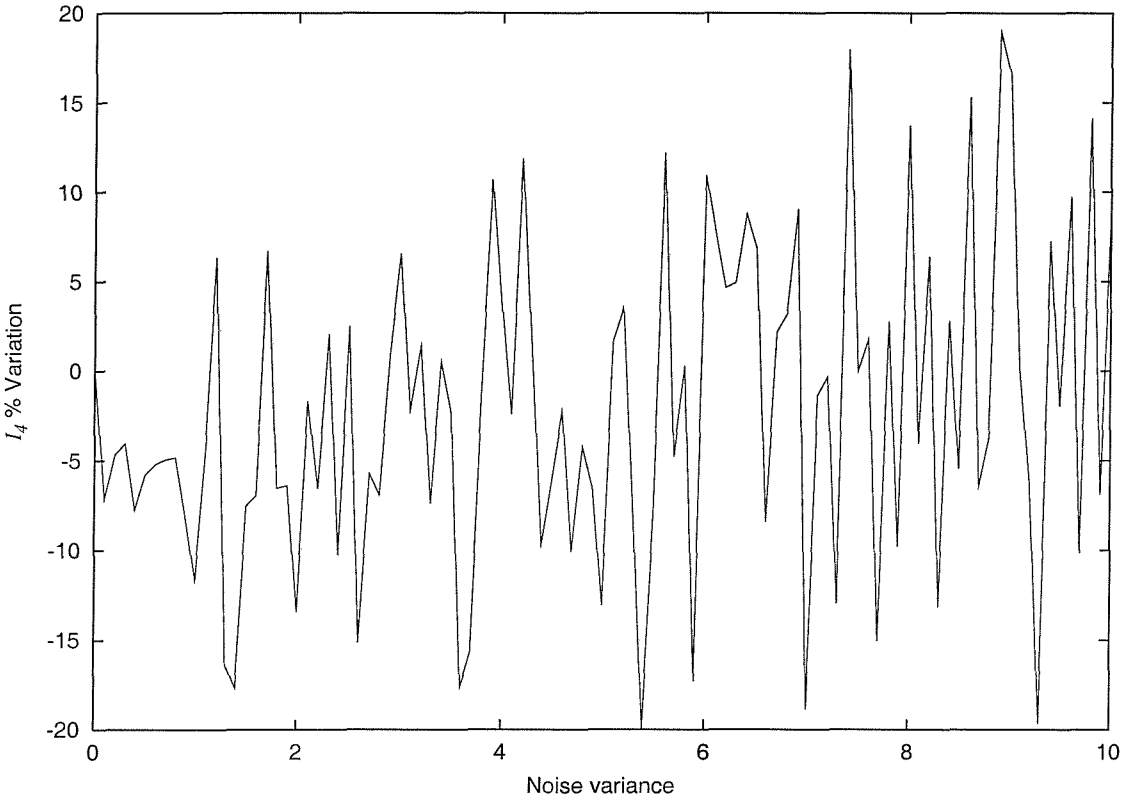


Figure A.10: Perimeter noise applied to the overlaid-shapes sequence - I_4

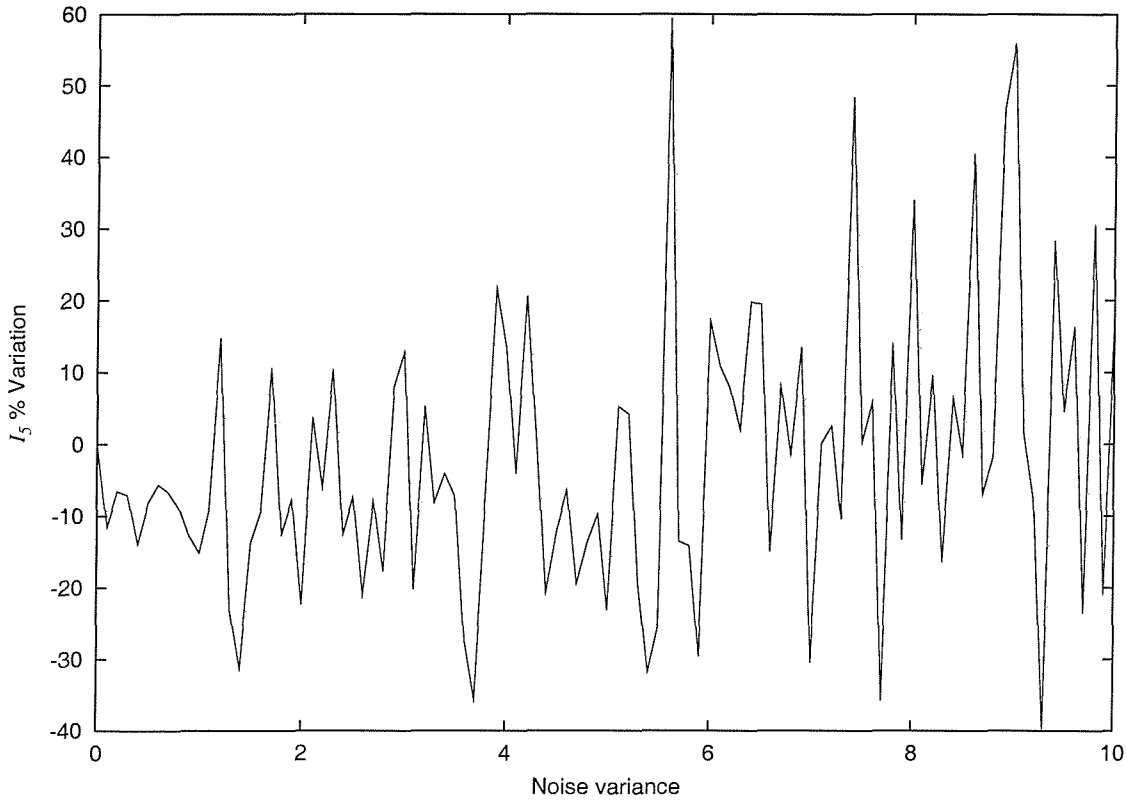


Figure A.11: Perimeter noise applied to the overlaid-shapes sequence - I_5

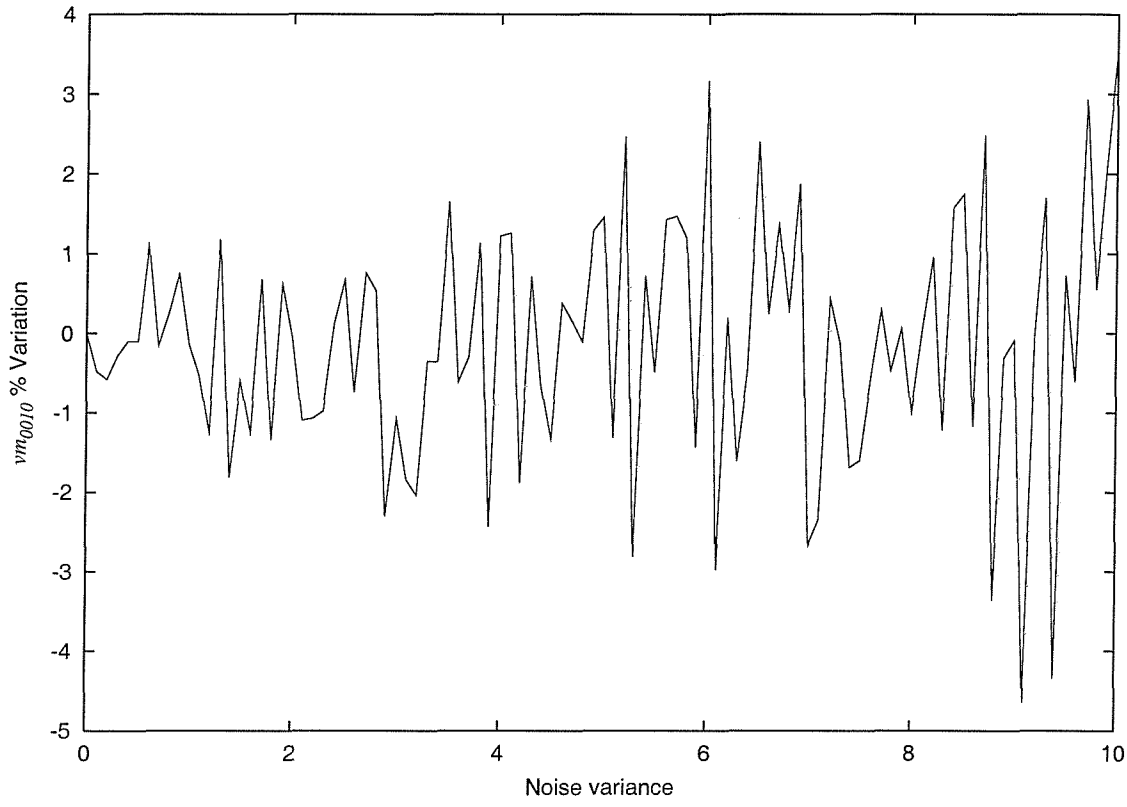


Figure A.12: Perimeter noise applied to the tug-boat sequence - vm_{0010}

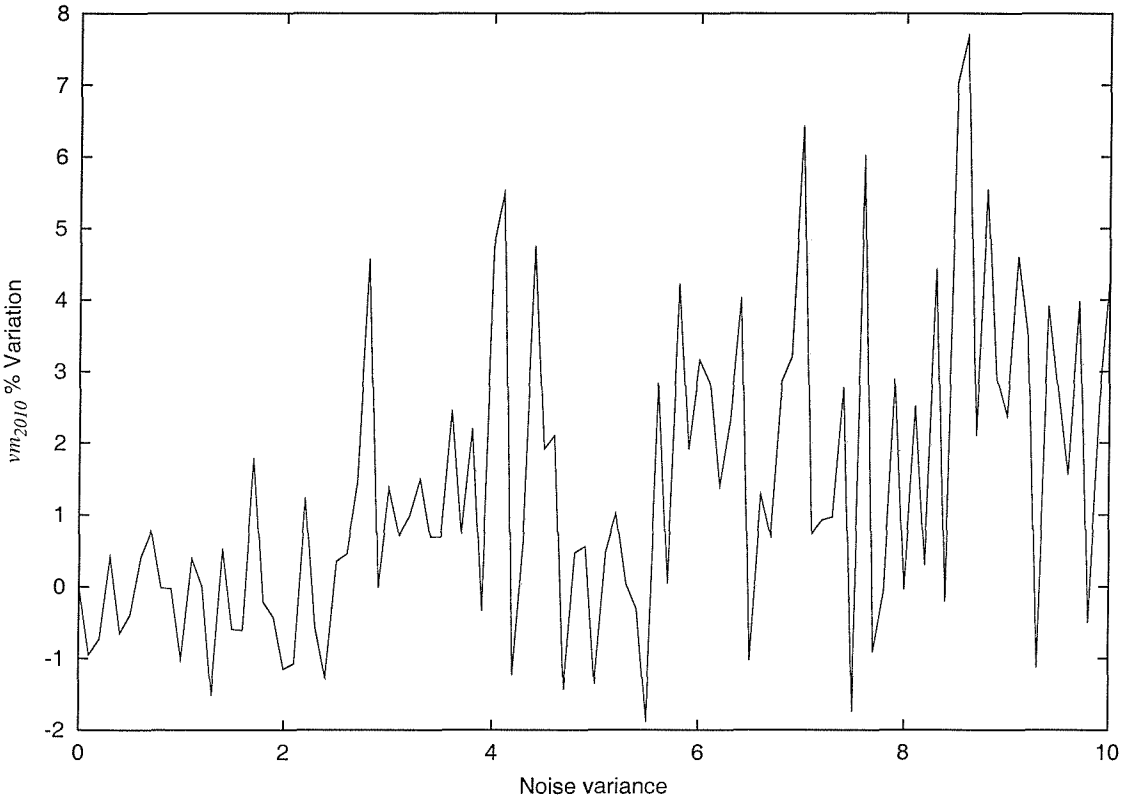


Figure A.13: Perimeter noise applied to the tug-boat sequence - vm_{2010}

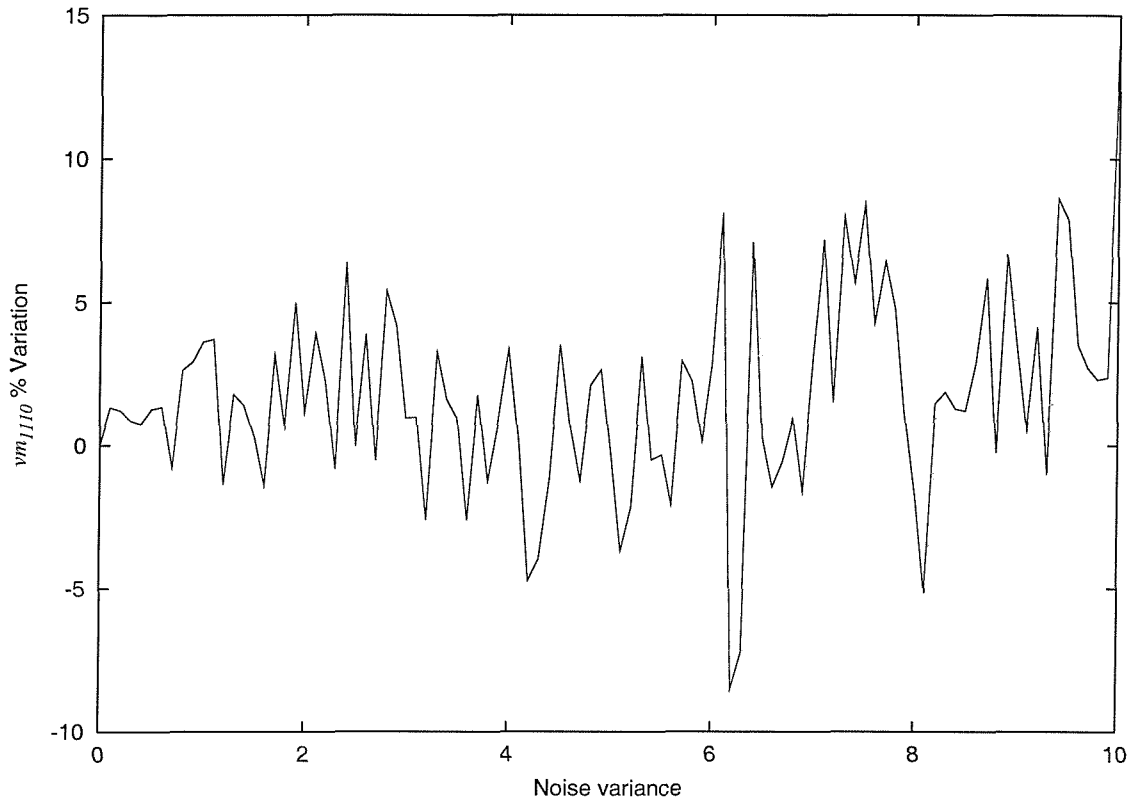


Figure A.14: Perimeter noise applied to the tug-boat sequence - vm_{2210}

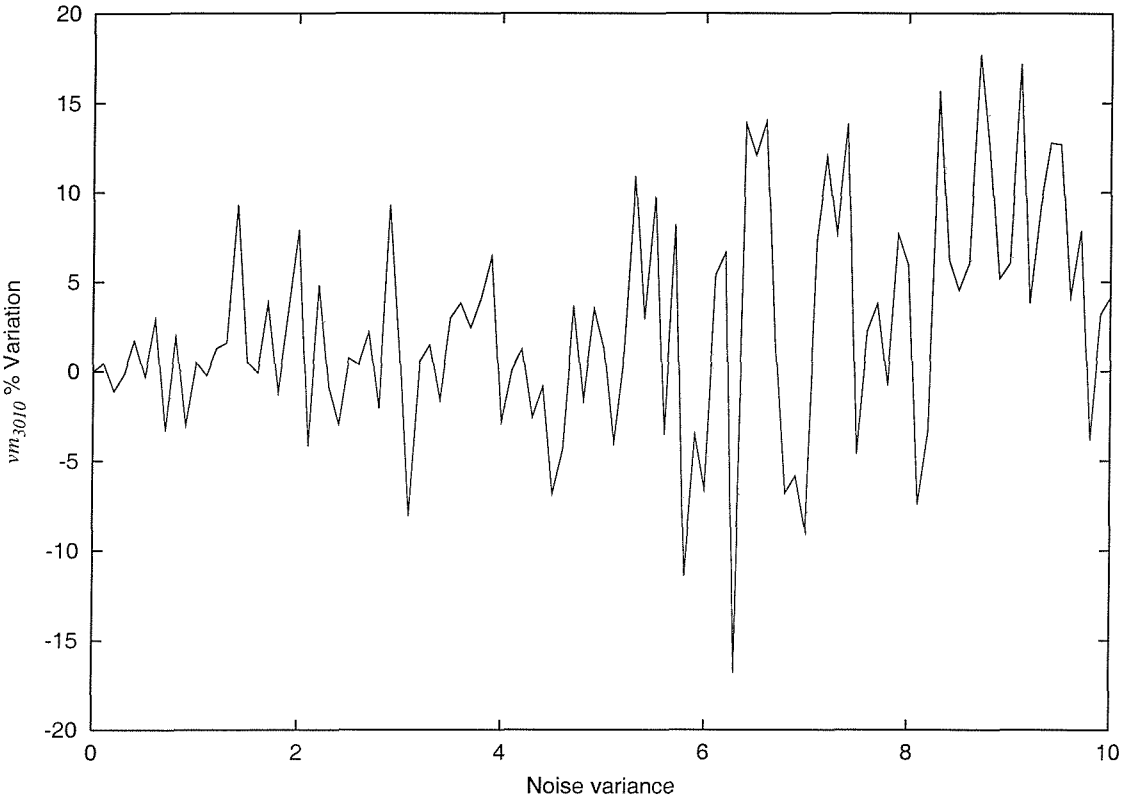


Figure A.15: Perimeter noise applied to the tug-boat sequence - vm_{1110}

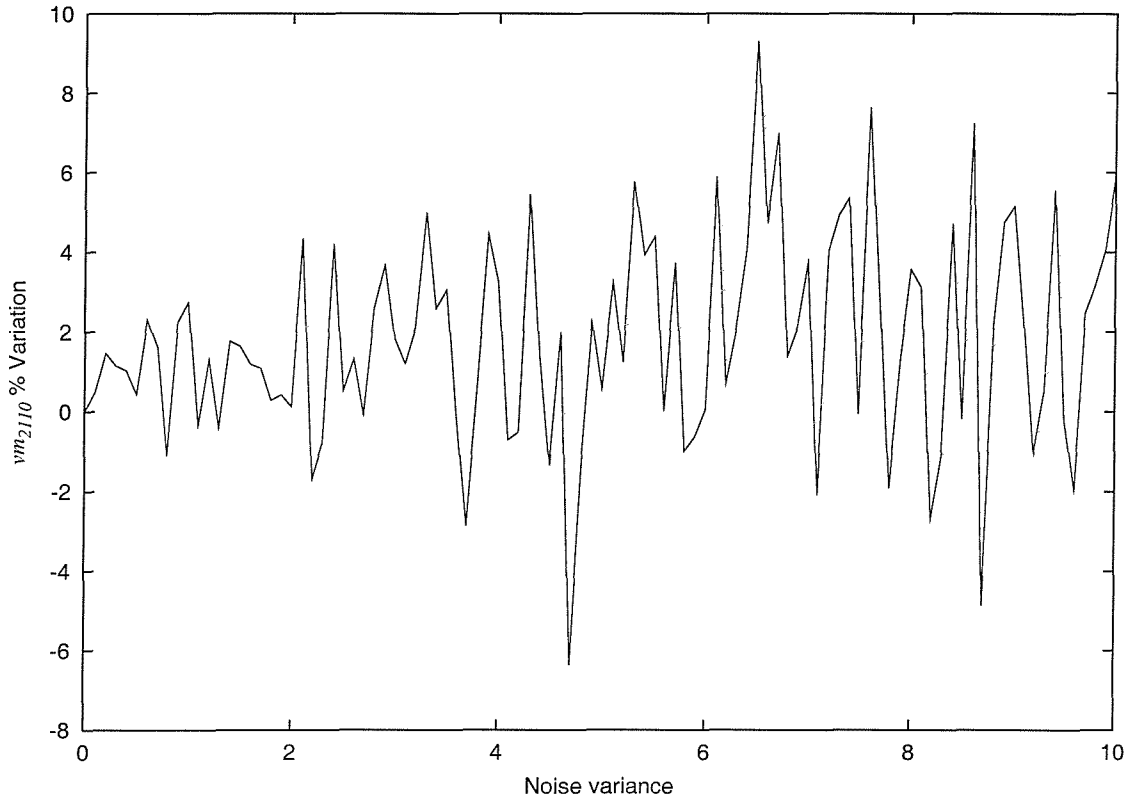


Figure A.16: Perimeter noise applied to the tug-boat sequence - vm_{2110}

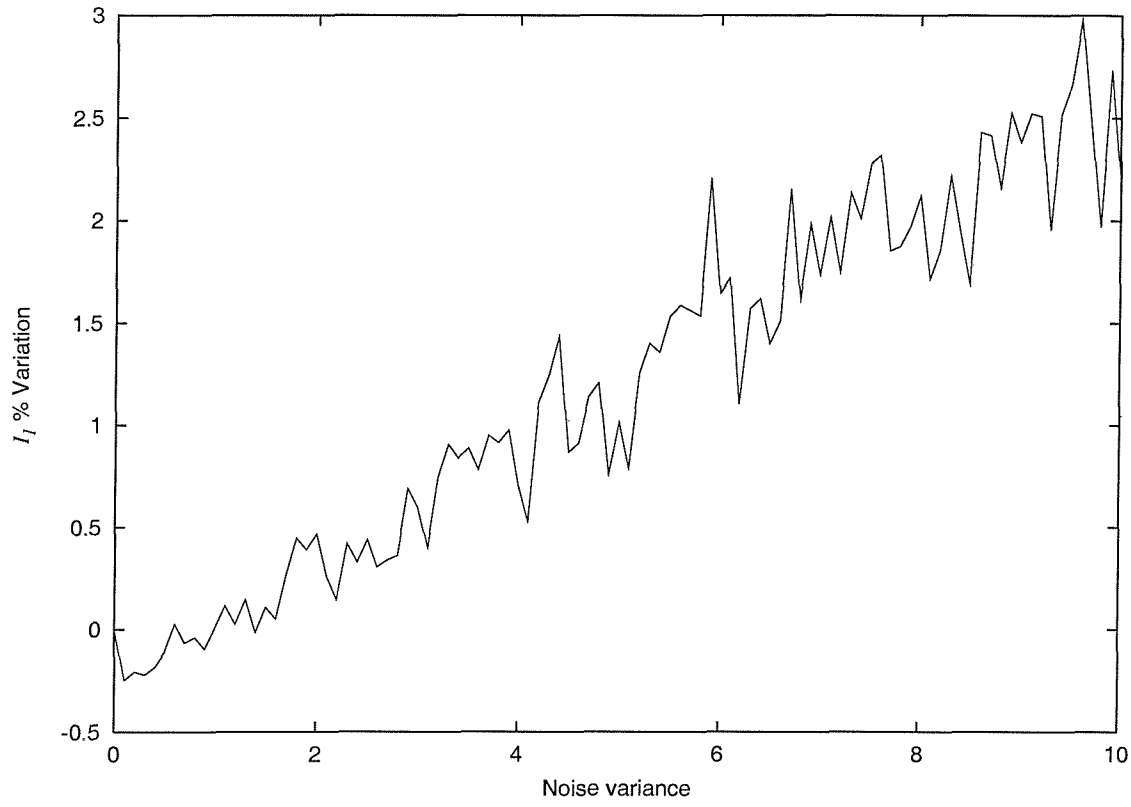


Figure A.17: Perimeter noise applied to the tug-boat sequence - I_1

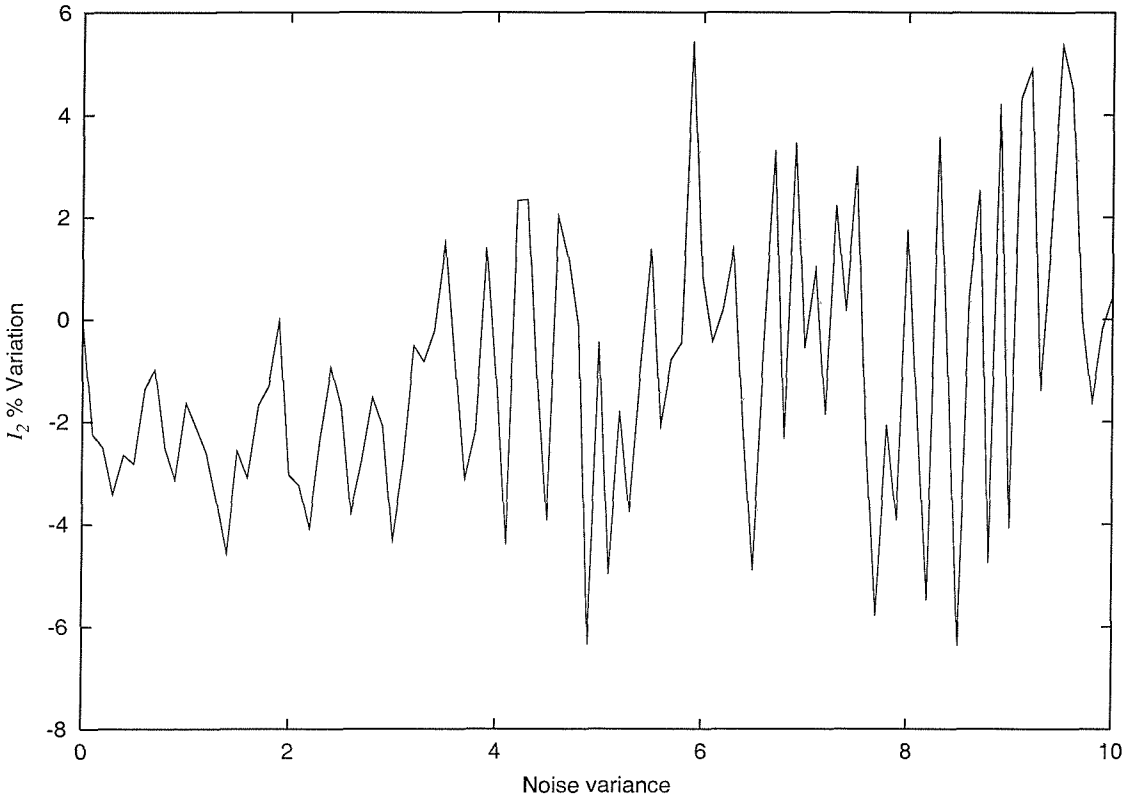


Figure A.18: Perimeter noise applied to the tug-boat sequence - I_2

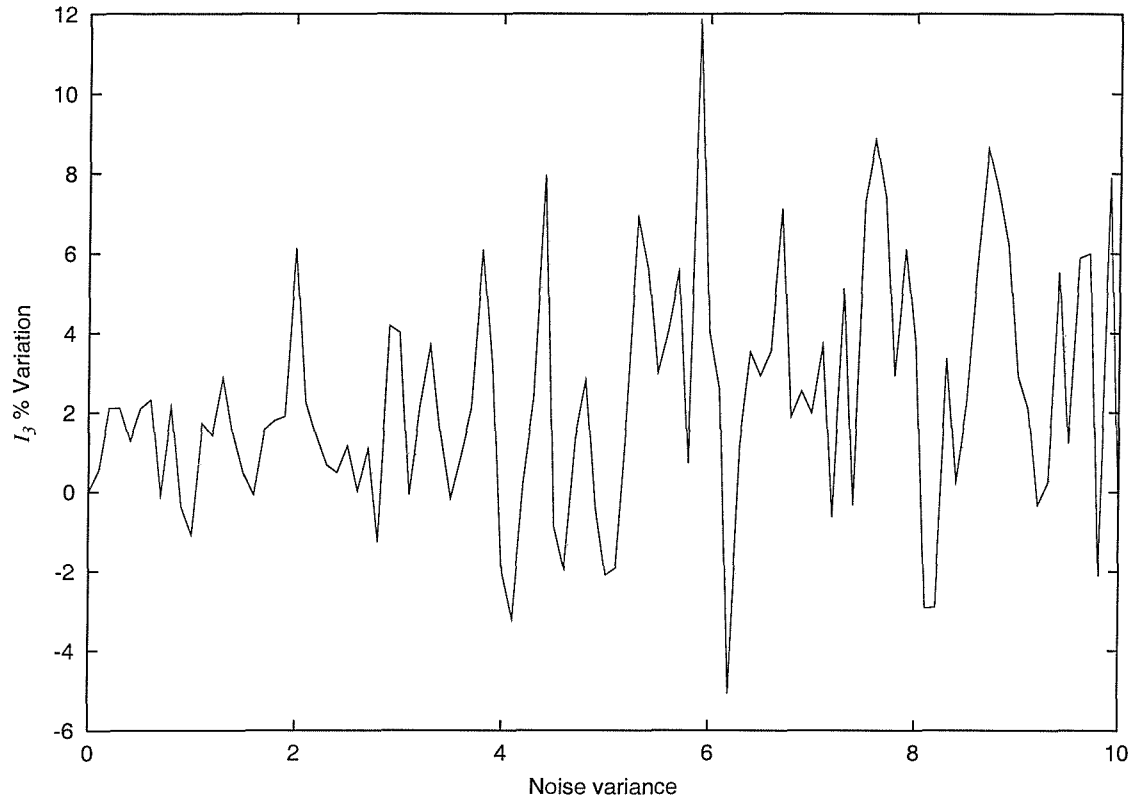


Figure A.19: Perimeter noise applied to the tug-boat sequence - I_3

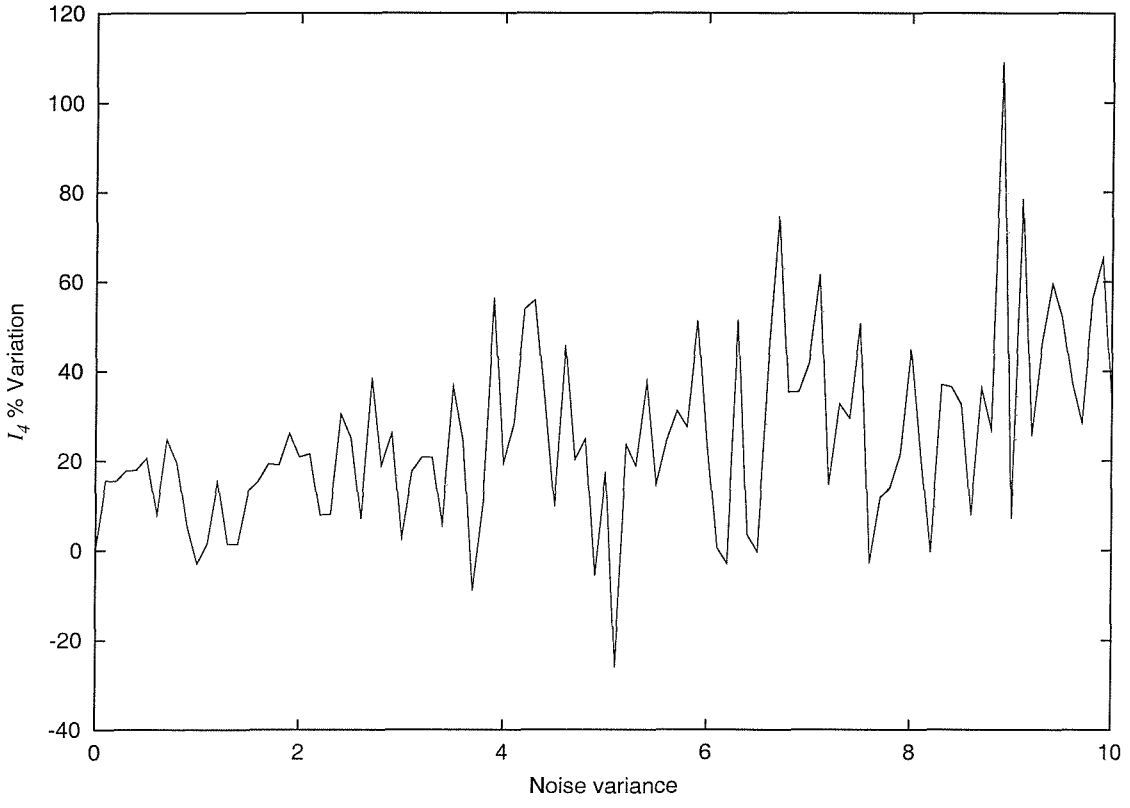


Figure A.20: Perimeter noise applied to the tug-boat sequence - I_4

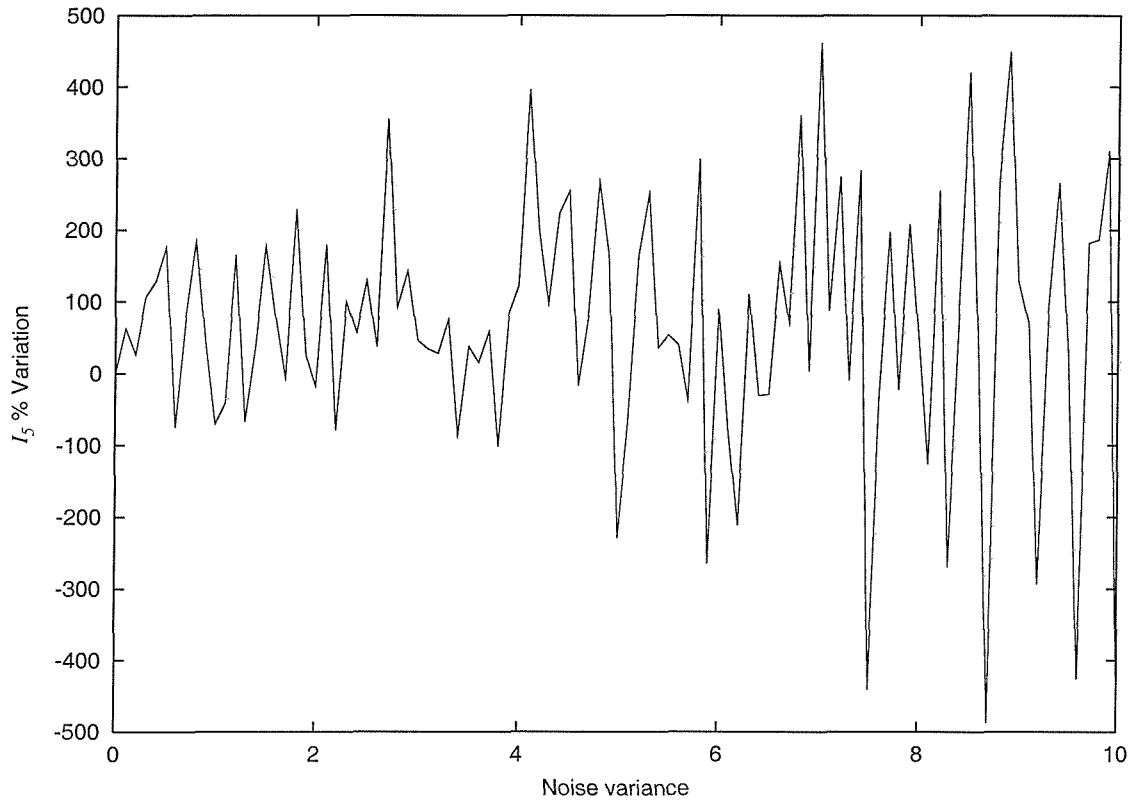


Figure A.21: Perimeter noise applied to the tug-boat sequence - I_5

Appendix B

Temporal statistical feature extraction

A statistical subject extraction method based on work by Jabri [32, 33] is detailed here. This offers considerably improved performance over simple basic foreground extraction operations, such as the subtraction of an image from the estimate of its background. The detection and extraction of moving humans is achieved in three parts. First of all a background model is needed, utilising this image background subtraction is then achieved. However, improvements over simple background subtraction are achieved by using both grey-scale and edge information. Here we are considering the algorithm's application to grey-scale images, however, it is worth noting that the method can be translated to colour images. This is achieved by applying everything detailed here separately to each of the three colour channels: red, green and blue, and then combining the results to form a colour image.

B.1 Edge data

Prior to generating the background model, edge data is needed. This is generated using a Sobel operator. The masks for the x and y edge directions respectively are:

$$\begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \qquad \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \tag{B.1}$$

These masks are then convolved with the image to produce the gradient images $G(x)$ and $G(y)$ for the x and y directions, respectively. (These results are then scaled to fit the image pixel range, here this was $0-255$.) This process is repeated for the whole sequence of images, resulting in 3 different source image sequences, the grey-level sequence (original data) and the two edge sequences.

B.2 Background model

The background model consists of mean images of both edge information and grey-scale. Together with these, the standard deviation images are calculated again for both the edge and grey-scale data. These images are used to aid the background subtraction detailed in the next section. The mean images are produced using a weighted sum technique, or exponential forgetting [22]. In this way the effects of changes in lighting and objects moving within the field of view can be removed. The background image effectively becomes a long term average of the scene, similar to a long exposure time on photographic film. The mean pixel value M_{xy} for I images is calculated using:

$$M_{xy} = \sum_{i=1}^{I} T\ P_{i_{xy}} + (1-T)\ M_{xy_{i-1}} \tag{B.2}$$

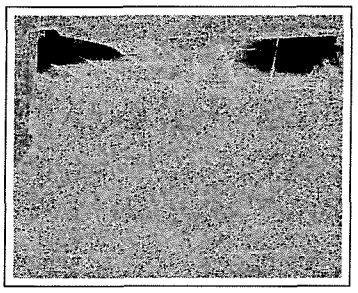
where $P_{i_{xy}}$ is the current pixel of image i, T is the time constant (or forgetting constant) and $M_{xy_{i-1}}$ is the mean pixel up to (and including) image $i-1$. The pixel variance σ^2_{xy} is also calculated using a weighted sum:

$$\sigma^2_{xy} = \sum_{i=1}^{I} T\ (P_{i_{xy}} - M_{xy})^2 + (1-T)\ \sigma^2_{xy_{i-1}} \tag{B.3}$$

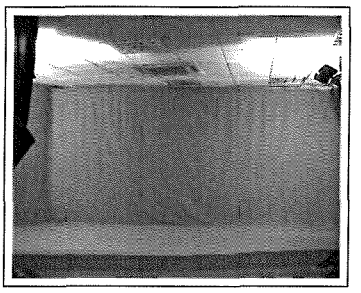
where $\sigma^2_{xy_{i-1}}$ is the pixel variance up to (and including) image $i-1$. The source images to create the background model can either be the sequence being extracted, or a sequence taken (preferably captured at the same time) containing just the background. The problem with the former is that the sequence must be long enough for the subject to have moved out of the area of interest, or ideally completely out of view. This is due to the nature of the weighted sum, if the subject is present in the last image used in Equations B.2 and B.3 then an area corresponding to their shape will appear in the final estimates of mean and variance. Example mean and variance images (constructed from a HiD background-only sequence) can be seen in Figures B.1a and b. For visual purposes only, the variance image has been histogram equalised to improve the contrast. The background model now consists of six different images. Three mean images of grey-scale and edge data ($M(grey)$, $M(edge\ x)$ and $M(edge\ y)$), and their three corresponding variance images ($\sigma^2(grey)$, $\sigma^2(edge\ x)$ and $\sigma^2(edge\ y)$).

B.3 Background subtraction

The subtraction is performed on both the grey-scale and edge data, the results of which are then combined. First we consider the grey-scale case. Rather than applying simple background subtraction on the grey level images, a confidence map is produced. This effectively labels regions within the image sequence as foreground



(a) Variance image



(b) Mean image

Figure B.1: Example background variance (histogram equalised) and mean images.

(i.e. moving objects or subjects) or background. The higher the confidence C, the more likely the grey-level pixel $P_{i_{xy}}$ is part of the foreground. The confidence level is set using two thresholding levels, m_g and n_g. These integer values set the confidence levels, i.e. how many standard deviations there are between the foreground and background objects. If the difference result $D_{i_{xy}} < n_g\sigma(grey)_{xy}$ then the pixel has a 0% confidence, whereas if $D_{i_{xy}} > m_g\sigma(grey)_{xy}$ then it has a 100% confidence level. For cases between these regions the grey-level confidence for an image i is calculated according to:

$$C(grey)_{i_{xy}} = \frac{(D_{i_{xy}} - n_g\ \sigma(grey)_{xy})}{(m_g\ \sigma(grey)_{xy} - n_g\ \sigma(grey)_{xy})} \tag{B.4}$$

while the difference $D_{i_{xy}}$ between the current image pixel $P_{i_{xy}}$ and the mean image M_{xy} determined by:

$$D_{i_{xy}} = \ \mid M(grey)_{xy} - P_{i_{xy}} \mid \tag{B.5}$$

These calculations result in a further image sequence of grey level confidence maps, an example image is shown in Figure B.2b. Subtraction of the edge data sequences, for image i is applied according to:

$$\triangle G(x)_{i_{xy}} = \ \mid M(edge\ x)_{xy} - G(x)_{i_{xy}} \mid \tag{B.6}$$

$$\triangle G(y)_{i_{xy}} = \ \mid M(edge\ y)_{xy} - G(y)_{i_{xy}} \mid \tag{B.7}$$

These two sequences are combined to give:

$$\triangle G_{i_{xy}} = \ \triangle G(x)_{i_{xy}} + \triangle G(y)_{i_{xy}} \tag{B.8}$$

which expresses the magnitude of the difference in both x and y. Next the edge reliability is calculated, this step aims to take into consideration the effects of noise on the edges. This produces measures which reflect how reliable the subtracted edge is, i.e. does the edge exist due to a lighting change, or is it present due to the subject moving. It is expressed as the ratio of the computed edge (difference) strength to

the confidence in the observed difference. The edge reliability R is defined as:

$$R_{i_{xy}} = \frac{\triangle G_{i_{xy}}}{G_{i_{xy}}} \tag{B.9}$$

the edge strength is expressed as:

$$G_{i_{xy}} = max\{G^c_{i_{xy}}, M(edge)_{xy}\} \tag{B.10}$$

and:

$$M(edge)_{xy} = \mid M(edge\ x)_{xy} \mid + \mid M(edge\ y)_{xy} \mid \tag{B.11}$$

$$G^c_{i_{xy}} = \mid G(x)^c_{i_{xy}} \mid + \mid G(y)^c_{i_{xy}} \mid \tag{B.12}$$

where G^c are the current image edge maps in x $(G(x)^c_{i_{xy}})$ and y $(G(y)^c_{i_{xy}})$. This means that in place of edge difference $D_{i_{xy}}$ used in Equation B.4, a measure of edge difference and reliability can be used, $R_{i_{xy}}\ \triangle G_{i_{xy}}$ producing edge confidence:

$$C(edge)_{i_{xy}} = \frac{(R_{i_{xy}}\ \triangle G_{i_{xy}} - n_e\ \sigma_{xy})}{(m_e\ \sigma_{xy} - n_e\ \sigma_{xy})} \tag{B.13}$$

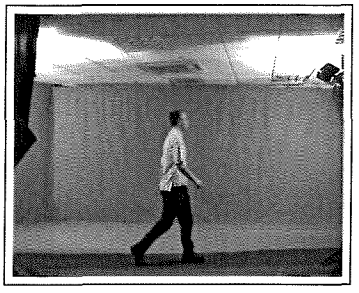
where m_e and n_e are the edge-system parameters which have similar properties to their grey-level counterparts, while σ_{xy} is the combined standard deviation for both the x and y direction edges $(\sigma(edge\ x) + \sigma(edge\ y))$. Figure B.2c shows an example edge confidence map. Finally, combining both confidence maps (edge and grey-scale) produces the combined confidence maps:

$$C_{i_{xy}} = max\{C(grey)_{i_{xy}}, C(edge)_{i_{xy}}\} \tag{B.14}$$

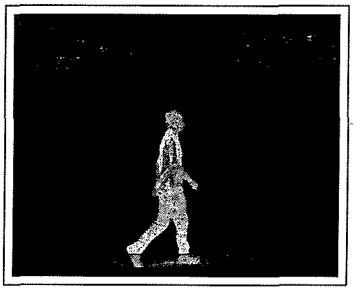
as shown in Figure B.2d. It is worth noting that by looking further at the edge phase information it is possible to classify the edges by type: occluding, occluded or background [32]. Care must also be taken to avoid divide-by-zero errors in Equations B.4, B.9 and B.13.

B.4 Locating and delineating the foreground

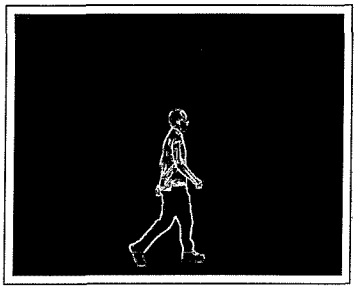
To improve the consistency of the combined confidence maps, median filtering is applied. A simple connected components algorithm is used, this enables objects to be filtered depending on their size. Hysteresis thresholding can be applied to remove any false positives not connected to a high confidence region. Finally, simple hole filling (i.e. holes within the subject's contour) is achieved using expanding and shrinking techniques [34] producing the final binary foreground confidence maps. Figure B.2e shows an example. Extraction of the subject can be achieved by logically ANDing the resultant confidence map image set with the original grey-scale



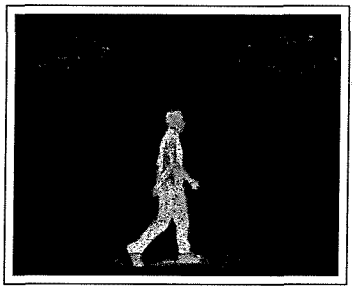
(a) Example image



(b) Grey-scale confidence



(c) Edge confidence



(d) Combined confidence

(e) Final confidence map

(f) Extracted subject

Figure B.2: An example image from a sequence showing it's colour subtracted version, edge confidence (histogram equalised) and the final confidence map.

images. An example result is shown in Figure B.2f. It is possible to further improve the confidence map by producing a contour around the main content of the map [32], i.e. the subject. By then extracting only the areas which are encompassed by this contour removes the possibility of holes appearing within the subject as the sequence progresses (holes that have been missed by the expanding/shrinking process).

Appendix C

Image re-sampling algorithm

This appendix details the algorithm used in Section 6.5 to re-sample an image to a lower resolution. It assumes that the original image is the highest resolution available and sub-pixel estimation is allowed, enabling any re-sampling size to be achieved. Figure C.2 shows an example satellite image of mount Fiji re-sampled to different resolutions using this algorithm. An implementation of the algorithm in 'C' for binary images can be found on the demonstration CD-ROM in Appendix D. If X_o and Y_o are the dimensions of the original image to be re-sampled (Figure C.1), then X_n and Y_n are expressed as:

$$X_n = fl\left(\frac{X_o}{\text{out}_{x_{max}}}\right)$$
$$Y_n = fl\left(\frac{Y_o}{\text{out}_{y_{max}}}\right) \tag{C.1}$$

$fl(x)$ is the value of x rounded down to the nearest integer (the 'floor' value). The opposite operation being rounding the value of x up to the nearest integer, expressed as the 'ceiling' value $cl(x)$. X_n and Y_n effectively define how many pixels of the original image, contribute to one pixel of the new image. $\text{out}_{x_{max}}$ and $\text{out}_{y_{max}}$ define the new re-sampling size, producing the area of the new image defined as:

$$\text{out}_x = 0 \rightarrow \text{out}_{x_{max}}$$
$$\text{out}_y = 0 \rightarrow \text{out}_{y_{max}} \tag{C.2}$$

The offset of the new sampling area from the origin of the original image ensures that the contents of the image do not move due to rounding errors (i.e. if the new pixel size is not an integer multiple of the original image size). This offset is defined as:

$$\text{offset}_x = \frac{mod(X_o\ ,\ \text{out}_{x_{max}})}{2}$$

$$\text{offset}_y = \frac{mod(Y_o\ ,\ \text{out}_{x_{max}})}{2} \tag{C.3}$$

where $mod(x, y)$ is the modulus operator, returning the remainder of x/y as X_n will be a rounded integer value due to dealing with discrete image coordinates. We can now define the new pixel size in terms of R and Q, Figure C.1. First we consider the case for point R. Its position is given by:

$$\begin{aligned} R_x &= (\text{out}_x\ X_n) + \text{offset}_x \\ R_y &= (\text{out}_y\ Y_n) + \text{offset}_y \end{aligned} \tag{C.4}$$

while its corresponding sub pixel values are:

$$\begin{aligned} R_{xmult} &= R_x - fl(R_x) \\ R_{ymult} &= R_y - fl(R_y) \end{aligned} \tag{C.5}$$

If $R_{xmult} > 0$ then the left (sub-pixel) edge exists, similarly if $R_{ymult} > 0$ then the top edge exists. Similarly, the coordinates for the point Q are defined by:

$$\begin{aligned} Q_x &= R_x + X_n \\ Q_y &= R_y + Y_n \end{aligned} \tag{C.6}$$

along with:

$$\begin{aligned} Q_{xmult} &= Q_x - fl(Q_x) \\ Q_{ymult} &= Q_y - fl(Q_y) \end{aligned} \tag{C.7}$$

If $Q_{xmult} > 0$ then the right edge exists, while if $Q_{ymult} > 0$ then the bottom edge exists. Q_{xmult} and R_{xmult} are dependent on the desired image re-sampling size. Depending on this re-sampling size, $Q_{xmult} \neq R_{xmult}$ is assumed and may be true. A similar assumption is valid for Q_{ymult} and R_{ymult}. We can now determine the contents of each new re-sampled pixel. Thus, the contributions of the inner block of the new pixel, the corners and the edge components are calculated. If the current pixel of position x, y of the original image is $P(x, y)$, then the area of the inner block I_b is defined by:

$$I_b = \sum_{x=cl(R_x)}^{fl(Q_x)} \sum_{y=cl(R_y)}^{fl(Q_y)} P(x, y) \tag{C.8}$$

The corners C_{1-4}, (starting at the top left and moving in a clock-wise direction) contribute through:

$$\begin{aligned} C_1 &= P(\ fl(R_x)\ ,\ fl(R_y)\ )\ R_{ymult}\ R_{xmult} \\ C_2 &= P(\ fl(Q_x)\ ,\ fl(R_y)\ )\ R_{ymult}\ Q_{xmult} \\ C_3 &= P(\ fl(Q_x)\ ,\ fl(Q_y)\ )\ Q_{ymult}\ Q_{xmult} \\ C_4 &= P(\ fl(R_x)\ ,\ fl(Q_y)\ )\ Q_{ymult}\ R_{xmult} \end{aligned} \tag{C.9}$$

Finally, considering the edge contributions E_{1-4}, the left and right edges (respectively) are:

$$\begin{aligned} E_1 &= \sum_{y=cl(R_y)}^{fl(Q_y)} P(fl(R_x)\ ,\ y)\ R_{xmult} \\ E_2 &= \sum_{y=cl(R_y)}^{fl(Q_y)} P(fl(Q_x)\ ,\ y)\ Q_{xmult} \end{aligned} \tag{C.10}$$

and the top and bottom edges respectively are given by:

$$\begin{aligned} E_3 &= \sum_{y=cl(R_x)}^{fl(Q_x)} P(x\ ,\ fl(R_y))\ R_{ymult} \\ E_4 &= \sum_{y=cl(R_x)}^{fl(Q_x)} P(x\ ,\ fl(Q_y))\ Q_{ymult} \end{aligned} \tag{C.11}$$

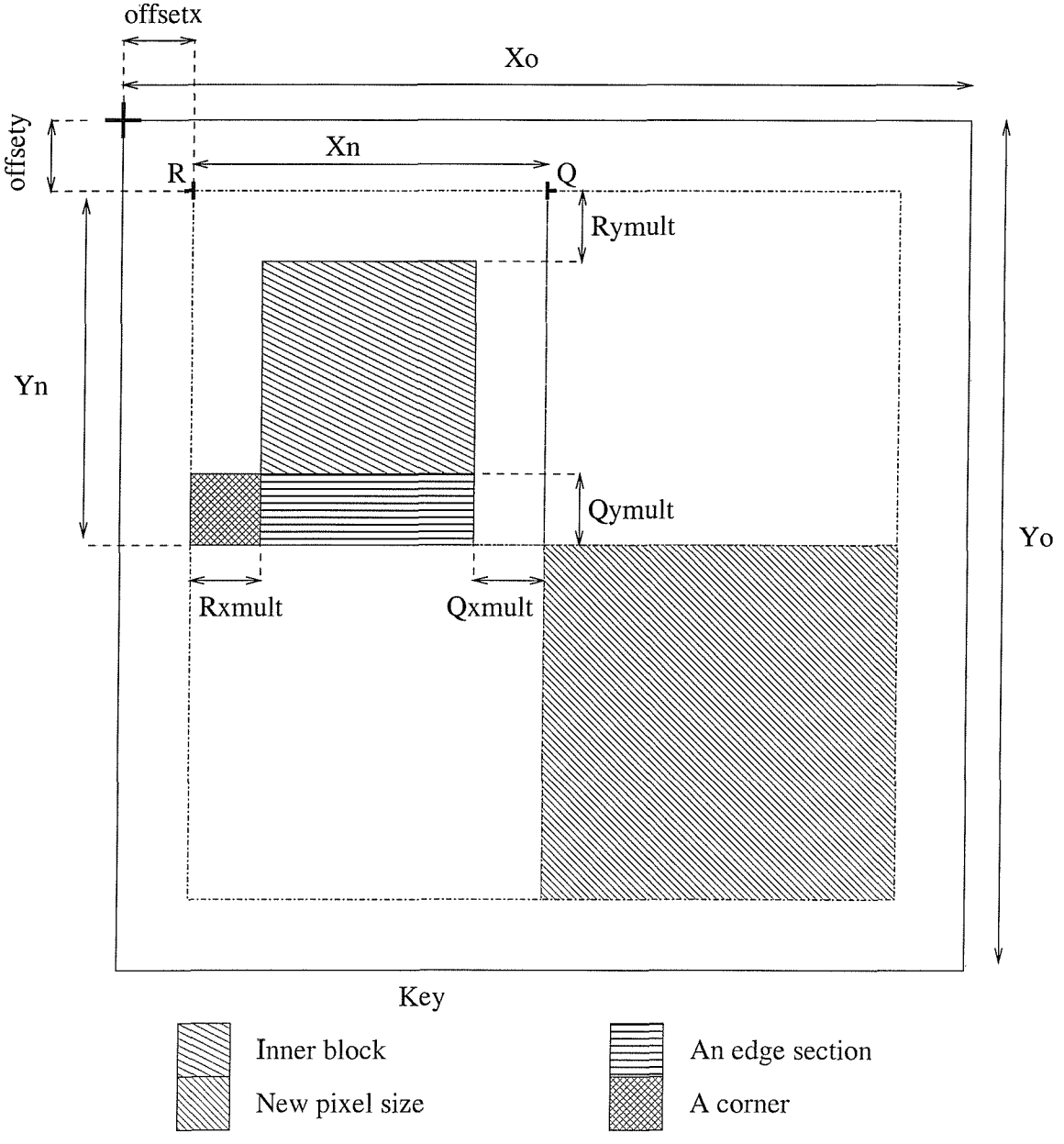


Figure C.1: Image re-sampling algorithm - defining the new pixel size.

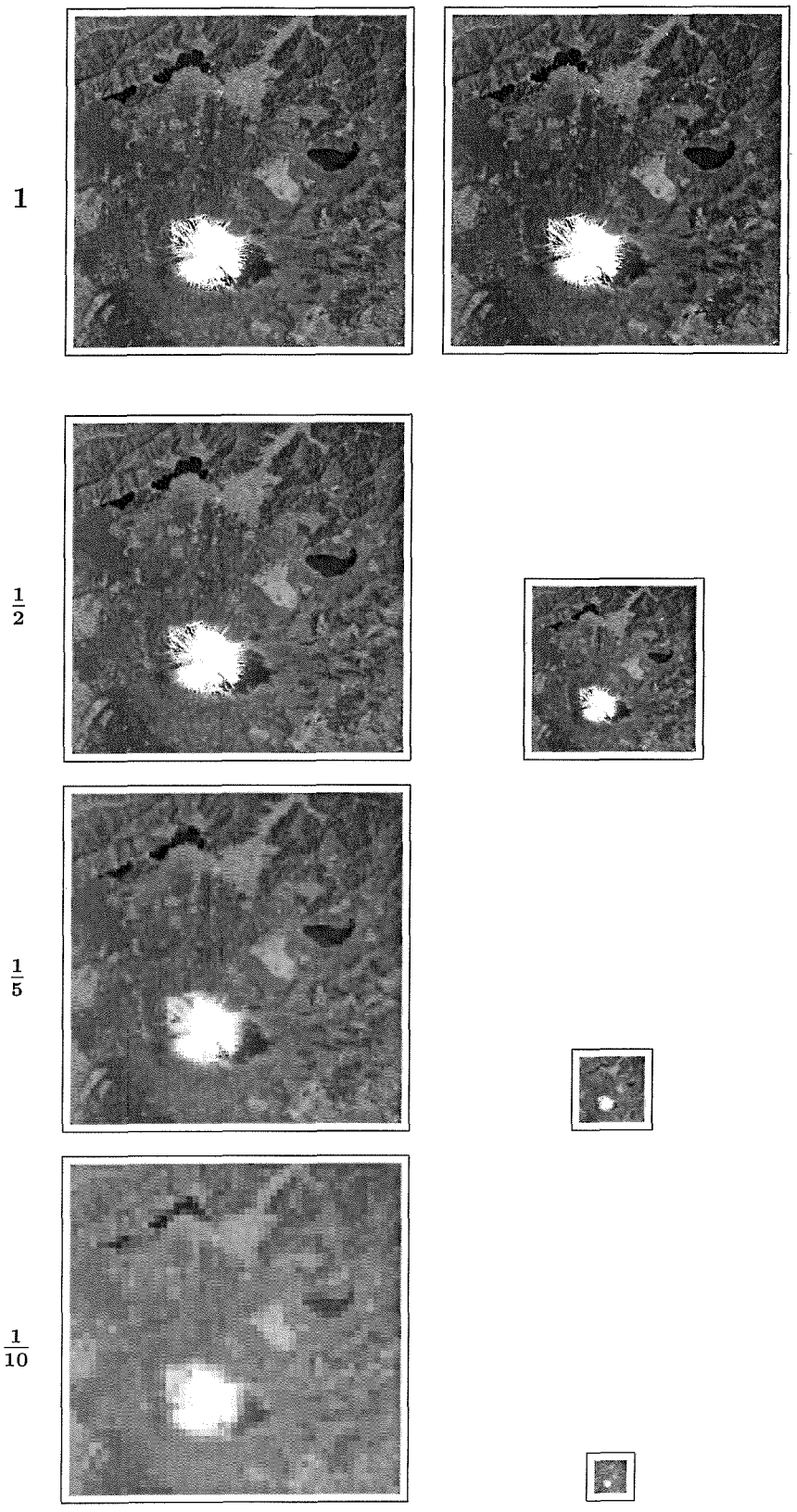


Figure C.2: Image re-sampling, original satellite image of mount Fuji at the top, (showing from left to right) the re-sampling scalar, resultant re-sampled image and their relative sizes.

Appendix D

Demonstration CD-ROM

A CD-ROM accompanies this thesis containing HTML pages describing and containing the following:

- Postscript and pdf versions of this thesis.
- Postscript versions of the author's publications.
- Animated sequences from each of the human gait databases.
- Animated extracted sequences from each of the human gait databases.
- Animated scatter plots of calculated features.
- Re-sampling algorithm 'C' source code.

These pages can also be found at the following website:
http://www.zepler.org/~jamie/velocitymoments.html

Further to this, tutorial information describing Zernike and velocity moments can be found at:
http://www.dai.ed.ac.uk/CVonline/LOCAL_COPIES/SHUTLER/CVonline.html