Extraction and Recognition of Periodically Deforming Objects by Continuous, Spatio-temporal Shape Description

S. D. Mowbray and M. S. Nixon

Electronics and Computer Science, University of Southampton,
Southampton, SO17 1BJ, UK

Abstract

We demonstrate a novel approach to modelling arbitrary temporally-deforming objects using spatio-temporal Fourier descriptors. This is a continuous boundary descriptor, which can handle shapes that vary in a periodic manner (such as a walking subject). As such, we can handle non-rigid, moving shapes that self-occlude. We show how this approach has led to successful shape extraction and description with both laboratory-sourced and real-world data. A consequence of exploiting temporal shape correlation in this approach has led to very good tolerance of noise and other positive performance factors. Further to this, our new approach holds sufficient descriptive power not only for extraction, but also for description purposes, and we have been pleased to note high recognition rates in human gait recognition on a large database.

1. Introduction

1.1 Finding Periodically-Deforming Objects

A prerequisite to being able to use an object's shape as means of recognizing it is the ability to successfully detect the particular shapes of interest in an image sequence. This is not a trivial problem and one which has been studied in the past by means of utilizing various forms of the Hough transform. In particular, there are two spatio-temporal versions of the Hough transform that extract shapes that move through time [13, 7]. Evidence-gathering techniques, such as these, *vote* for free parameters of the shape model, typically the shape's location in a Cartesian space, although other parameters such as scale and orientation may also be sought.

Previous approaches to finding moving shapes have only dealt with static shapes, whereas many shapes in the real-world, especially the shapes of live subjects, deform in a periodic manner as they move. Through the use of spatio-temporal Fourier descriptors, a new variant of the Hough transform is developed, which can extract moving, deformable objects from an image sequence (such as walking human subjects). This new evidence-gathering technique not only proves to inherit many desirable features from the standard Hough transform, such as its ability to deal with noise and occlusion, but also provides a continuous, spatio-temporal shape representation to deal with the effects of discretisation, and the computational benefit of having a parameter space with a fixed number of dimensions, whilst still being capable of describing arbitrary shapes, shape deformation, and motion.

1.2 Gait as a Biometric

Recently, a significant amount of attention has been devoted to the use of human gait patterns as a biometric and to the analysis of human motion in general. Several models have been proposed for the description of the human body and also for the description of human gait [1, 6]. Human gait has many advantages over other biometrics, but perhaps its two most notable advantages are that it is non-invasive – offering a means to verify identity without a subject's active participation, and that it offers a means of offering recognition at a distance.

Many descriptors of human gait are kinematic in nature, relying on geometric descriptions of the various body parts and mathematical modelling of the affine transformations which describe their movement. Many kinematic *features* exist, such as the angles of various body parts through time, their velocity, and their acceleration. Other studies have taken a more statistical approach to produce a unique gait descriptor. These have included the use of Principal Component Analysis (for dimensionality compression) combined with Cananonical Analysis (for classification) [9], the use of velocity moments [15], and the use of Hidden Markov Models [11]. Other approaches have included the use of area-based features [4], and the use of static parameters, such as height and stride-length [10]. Research has also been carried out into describing human motion by analyzing the bilateral symmetry inherent in human gait [2, 8].

0-7695-2158-4/04 $20.00 (C) 2004 IEEE

This study demonstrates a new, model-based approach that captures the full movement and deformation of the body, rather than just a specific body part, and by utilizing only the boundary information of the object.

2 Spatio-Temporal Fourier Descriptors

Two-dimensional shapes are usually represented on a Cartesian plane, where the x and y co-ordinates of the shape can be thought of as a function of the shape's arc-length index, l. If we define a mapping from the Cartesian plane to a complex plane, however, then we can represent a shape's boundary as a complex function of arc-length:

$$c(l) = x(l) + j.y(l) \tag{1}$$

Due to the periodicity of $c(l)$ it is possible to represent the shape's boundary using a Fourier series, with the coefficients of the series, a_{xk}, b_{xk}, a_{yk}, and b_{yk}, being the Fourier descriptors of the shape:

$$\begin{aligned} c(l) = \quad & \tfrac{a_{x0}}{2} + \int\limits_{k=1}^{\infty} a_{xk} \cos\left(\tfrac{kl2\pi}{L}\right) + b_{xk} \sin\left(\tfrac{kl2\pi}{L}\right) dk + \\ & j\left(\tfrac{a_{y0}}{2} + \int\limits_{k=1}^{\infty} a_{yk} \cos\left(\tfrac{kl2\pi}{L}\right) + b_{yk} \sin\left(\tfrac{kl2\pi}{L}\right) dk\right) \end{aligned} \tag{2}$$

If we now consider a shape which deforms between t and T in time, we can model the periodicity of the deforming shape s, at arc-length index l as

$$s(t, l) = s(t + T, l) \tag{3}$$

Given this, it is possible to model the whole periodically deforming boundary of a shape in this shape sequence as a two-dimensional complex Fourier series:

$$s(t,\ l) = \int\limits_{k_t=0}^{\infty} \int\limits_{k_l=0}^{\infty} \hat{s}(k_t, k_l) e^{j2\pi\left(\frac{t.k_t}{T} + \frac{l.k_l}{L}\right)}\ dk_l\ dk_t \tag{4}$$

given here in complex form for brevity. The Fourier coefficients of this series characterize one whole period of the shape's movement.

If we consider the discrete case for a periodically deforming and moving shape then T is equivalent to the number of images in one period of motion, L represents the length of the boundary of the object, indexed by l, and the Fourier coefficients, $\hat{s}$, can be calculated using a discrete, two-dimensional, complex Fourier transform.

$$\hat{s}(k_t,\ k_l) = \frac{1}{T.L} \sum_{t=0}^{T} \sum_{l=0}^{L} s(t,\ l) e^{-j2\pi\left(\frac{k_t.t}{T} + \frac{k_l.l}{L}\right)} \tag{5}$$

We have earlier investigated the descriptive properties of spatio-temporal Fourier descriptors [12] and now show that we can also use them as a basis to formulate a new extraction technique with the capability of extracting periodically-deforming shapes from real-world imagery.

3. The Continuous Deformable Hough Transform

3.1 Background

Many image sequences contain a significant amount of temporal correlation, a fact which is frequently utilized by video compression techniques. The Velocity Hough Transform (VHT)[13] was the first evidence-gathering technique to use this correlation to extract the optimal parameters of a linearly moving conic section. Simple extensions to the VHT made it possible to extract any shape and motion combination, given that both were able to be modelled parametrically.

The VHT was originally developed as an extension of the Hough Transform for circles by adding a velocity parameter into the formulae used to cast votes into the accumulator space. The original VHT extended the Hough transform for circles to include velocity, as follows:

$$\begin{aligned} a_x &= c_x + r.\cos(\theta) + v_x.t \\ a_y &= c_y + r.\sin(\theta) + v_y.t \end{aligned} \tag{6}$$

where a_x and a_y are the coordinates of the vote to be cast, c_x and c_y are the center coordinates of the circle to be cast into the accumulator, r and θ are the polar parameters of the circle in question, v_x and v_y describe the linear velocity of the circle and t represents a time reference relative to the start of the object's motion.

During the voting process, the time reference, t, and the x-axis and y-axis velocities, v_x and v_y, are used to determine the location of the center coordinates of the circle in the initial frame of motion, at $t = 0$, therefore focusing all votes for the correct circle onto one set of a_x and a_y coordinates. Thus, the coordinates voted for when using the VHT are the center coordinates $(a_x,\ a_y)$ of the circle at its initial time reference with a linear velocity described by $(v_x,\ v_y)$.

The VHT was extended[7] to include a continuous shape model, using Elliptic Fourier Descriptors, to create the Continuous Velocity Hough Transform (CVHT). Using a continuous shape model ensures that discretisation errors, usually associated with applying affine transformations to discretized shape models, can be minimized, and also allows for undetermined points in a shape model to be interpolated with ease.

All previous approaches to shape detection through evidence-gathering have dealt with rigid shapes. We now describe a new form of Hough transform, the Continuous

Deformable Hough Transform, which is specifically designed to deal with non-rigid, moving shapes.

3.2 Theory of The Continuous Deformable Hough Transform

If we represent a spatio-temporal shape sequence, as described in section 2, as a complex Fourier descriptor matrix, $\hat{s}(k_t,\ k_l)$, then we can define a kernel that defines the shape of votes to be cast in the accumulator space for each feature point in a test image sequence (each edge pixel). This kernel is a combination of shape sequences, at various spatial, temporal, and motion scales.

The basis for the CDHT kernel is a normalized shape sequence descriptor, derived from Eq. 5. This is the complex spatio-temporal Fourier descriptor matrix of the shape we wish to find in the test sequence. The shape sequence descriptor is normalized so that the shape's maximum height and width with respect to the whole image sequence are set to unity. One drawback to this normalization approach is that the aspect ratio of the shape is lost. There are reasons for this, however, particularly when the technique is applied to finding humans, as humans themselves do not have a fixed aspect ratio, and thus allowing non-uniform scaling of both width and height (and therefore justifying normalization of both) is applicable here.

The motion model, provided by the spatial DC components, is also normalized independently of shape size so that the linear distance that the shape travels (in its direction of motion) falls in the range $[0, 1)$, with the shape's initial center of mass at the beginning of the shape sequence being $(0, 0)$. An example of a reconstructed normalized shape sequence descriptor is shown in Figure 1.

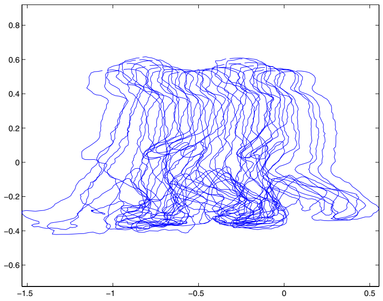


Figure 1: Example normalized shape sequence descriptor

During the following formal definition of the kernel of the CDHT, the normalized shape sequence descriptor matrix, as described above, will be referred to as the Shape Sequence Template (SST). This is simply a matrix of spatio-temporal descriptors from which the algorithm can reconstruct a normalized trace of a shape sequence.

The CDHT kernel, $\bar{\omega}$, can thus be defined as being the rescaled version of a SST, for all scale values we wish to search for, transformed into the time domain:

$$\bar{\omega}(t, l, x_s, y_s, t_s, v_s) = \int_{k_t=0}^{K_T-1} \left[\int_{k_l=0}^{K_L-1} SST(k_t, k_l, x_s, y_s, t_s, v_s) e^{\frac{j2\pi k_l l}{K_L}} dk_l \right] e^{\frac{j2\pi k_t t}{K_T}} dk_t \tag{7}$$

where $SST(k_t, k_l, x_s, y_s, t_s, v_s)$ is a scaling function which scales the SST appropriately for the spatial scaling variables, x_s and y_s, the temporal length of the sequence (in frames), t_s, and the 'velocity' scale factor, v_s. The separate velocity scaling factor is required here, as we should not assume a relationship between shape-size and velocity, or sequence-length and velocity.

3.3 The CDHT Voting Process

The kernel defines the shape sequence of votes to be cast into the accumulator for each feature point (edge pixel) in a test image sequence. This is a combination of shape sequences at varying spatial, temporal, and velocity scales.

Each sequence of votes is cast into the accumulator by offsetting it from the co-ordinates of each feature point in the test image sequence, IS, defined by

$$IS = \left\{ \bar{\lambda}(f, \mathbf{p}) \mid f \in D_f,\ \mathbf{p} \in D_p \right\} \tag{8}$$

where $\bar{\lambda}(f, \mathbf{p})$ is a function that defines the feature points, $\mathbf{p}$ with co-ordinates (p_x, p_y), in an image for each frame, f, in an image sequence, and D_p and D_f define the domains of pixels in an image and frames in an image sequence respectively.

Given this, the accumulator vote function is defined as

$$A_{fp} = \left\{ \begin{array}{l} \bar{\lambda}(f, p_x) - \Re\left\{ \bar{\omega}(f - t, l, x_s, y_s, t_s, v_s) \right\}, \\ \bar{\lambda}(f, p_y) - \Im\left\{ \bar{\omega}(f - t, l, x_s, y_s, t_s, v_s) \right\} \end{array} \right. \tag{9}$$

where $t \in D_t, l \in D_l, f \in D_f, \mathbf{p} \in D_p$ for $f \geq t$ where D_t is the domain of the possible temporal locations for the start of the shape sequence and D_l is the domain of the arc-length parameter of each shape. A_{fp} then defines a set of vote coordinates for which votes will be cast in the accumulator. It is necessary to decompose the kernel here into its real and imaginary parts, which determine the offsets along the x-axis and y-axis respectively of the feature point from the location of the vote in the accumulator.

With the accumulator vote function now defined, it is now necessary to define a matching function that will map the vote coordinates in set A_{fp} into the accumulator space. This matching function determines if a point in the accumulator's parameter space $\boldsymbol{\alpha}$, should be incremented for a point, $\mathbf{a}$, in set A_{fp}. The simplest form of matching function simply increments a matched accumulator point by unity:

$$M(\boldsymbol{\alpha}, \mathbf{a}) = \begin{cases} 1 & \boldsymbol{\alpha} = \mathbf{a} \\ 0 & \boldsymbol{\alpha} \neq \mathbf{a} \end{cases} \tag{10}$$

This matching function is then applied to A_{fp} for a range of parameter values, thus fully defining the continuous version of the CDHT as

$$\begin{aligned} &CDHT(\boldsymbol{\delta}, t, x_s, y_s, t_s, v_s) = \\ &\int_l \int_{t=0}^{F} \int_{f=t}^{F} \int_p M\Big((\boldsymbol{\delta}, t),\ \bar{\lambda}(f, \mathbf{p}) - \boldsymbol{\omega}\Big)\, dp\, df\, dt\, dl \end{aligned} \tag{11}$$

where

$$\boldsymbol{\omega} \equiv \left(\begin{array}{c} \Re\left\{\bar{\omega}(f - t, l, x_s, y_s, t_s, v_s)\right\}, \\ \Im\left\{\bar{\omega}(f - t, l, x_s, y_s, t_s, v_s)\right\} \end{array} \right) \tag{12}$$

In Eq. 11, $\boldsymbol{\delta}$ is the spatial translation vector and t is the temporal translation (in frames) to the center of mass of the first shape in the shape sequence. This continuous parameter space is then sampled into a discrete parameter space, given by

$$\begin{aligned} &DCDHT(\boldsymbol{\delta}, t, x_s, y_s, t_s, v_s) = \\ &\sum_{l \in D_l} \sum_{t=0}^{F} \sum_{f=t}^{F} \sum_{p \in D_p} M\Big((\boldsymbol{\delta}, t),\ \bar{\lambda}(f, \mathbf{p}) - \boldsymbol{\omega}\Big) \end{aligned} \tag{13}$$

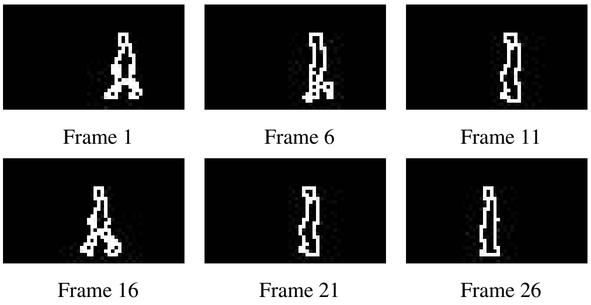
with the global maximum of this space being indexed by the estimated parameters of the shape-sequence model appearing in the image sequence.

3.4 Noise Analysis

The Hough transform and its derivatives are well known to be very robust to noise due to their implicit evidence-gathering properties. Noise tends to lead to false feature points in the image sequence, and as such, false loci of votes in the accumulator of the Hough transform. Conversely, however, any true feature points (edge pixels of the actual shape sequence) remaining after the addition of noise will also cast loci of votes in the accumulator, which should coincide to provide a global maximum indexed by the correct parameters of the shape. So long as the global maximum of the accumulator is large enough to not be masked by the noise then the correct shape sequence parameters will be found.

The image sequence used in the noise performance tests was derived from the Fourier descriptor-based shape sequence template of a real-world image sequence. Using this shape sequence template the test sequence was reconstructed at a low resolution in order to make the computational demands of the performance tests detailed here feasible. Although the test image sequence is not a true synthetic image sequence, as it is generated from imagery of the real-world, it can be considered to be synthetic as there is no background noise, with the only feature pixels being from the shape sequence itself. Edge detection is obviously not necessary (and perhaps not desirable) on a synthetic image sequence such as this, as all we wish to compare is the effect of noise, rather than the performance of an edge detector at removing noise and detecting true edges. All real-world image sequences would obviously need to be edge detected prior to being used as input to the CDHT.

Selected frames from the test image sequence are shown in Figure 2



Frame 1 Frame 6 Frame 11

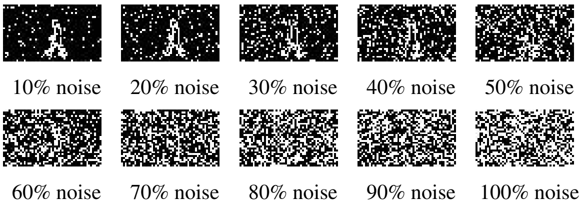
Frame 16 Frame 21 Frame 26

Figure 2: Synthetic Shape Sequence

The CDHT was tested for its robustness to noise using 11 noise levels, from 0% – 100%. An effective and adequate noise model for binary synthetic edge data (such as that used here) is to simply add false edge pixels to the background and false background pixels to true edges with equal probability (where a background pixel is assumed to be 0 and an edge pixel a 1), in effect, a form of salt and pepper noise. If we presume no previous knowledge of the image then we can assume each pixel has an equal probability of being a background pixel or an edge pixel and therefore to add noise we can merely assign a pixel in the image to be either a background pixel or an edge pixel, with equal probability. For example, adding 10% noise would change roughly 5% of the background pixels into edge pixels and roughly 5% of the edge pixels into background pixels. At 100% noise, the image would contain a purely random distribution of edge and background pixels, with any true background or edge pixels remaining only by chance. The varying noise levels used during testing, added to one frame from the test sequence, are shown in Figure 3.

The result of the noise performance test using a set of fifty noisy sequences corrupted by varying levels of random noise is shown in Figure 4.

These results demonstrates that the CDHT has an excellent tolerance to noise. It should be noted that, to the human eye, the apparent motion of a human is not visible at noise levels of 60% and above. The CDHT, however, still detects the correct location of the human in the test image sequence,



10% noise 20% noise 30% noise 40% noise 50% noise

60% noise 70% noise 80% noise 90% noise 100% noise

Figure 3: Frame 16 with varying noise levels added

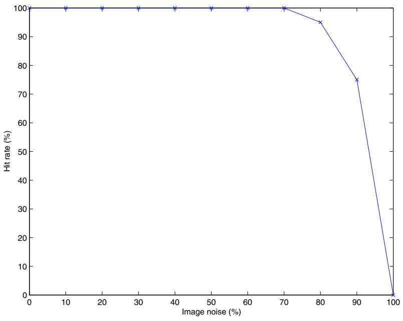
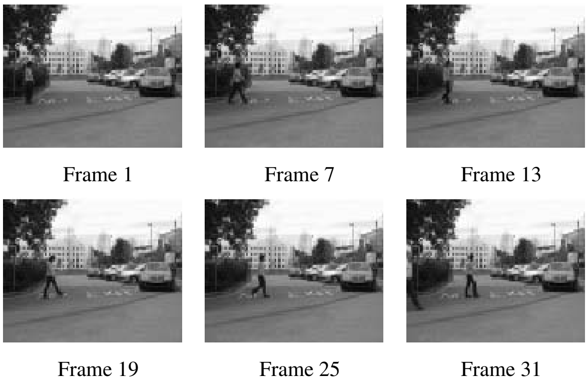


Figure 4: Noise performance

both spatially and temporally at noise levels of up to 80%. The ability to tolerate such extreme levels of noise is due to the global spatio-temporal evidence gathering nature of the CDHT. For example, if one frame in the CDHT is corrupted to a point where it is no longer recognizable, then this does not have such a great impact on the final result, whereas other evidence gathering techniques, such as the GHT are limited to the amount of information contained in one frame of an image sequence, which may prove to be insufficient to track the object correctly if the object becomes occluded.

3.5 Real-world image data

The above results indicate that the CDHT works well when given a synthetic input, but in practice edge detected images sequences are never so well defined. Noise is usually always present in an image to some extent, and even the best edge detection techniques result in false edges being detected while true edges are not. Due to this, the CDHT was tested on the real-world image sequence, selected frames of which are shown in Figure 5.

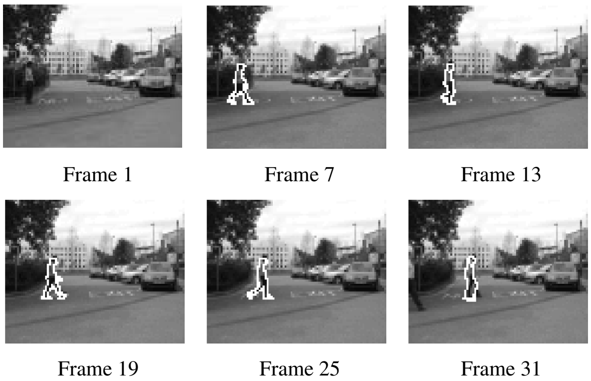


Frame 1 Frame 7 Frame 13

Frame 19 Frame 25 Frame 31

Figure 5: Example real-world image sequence

The image sequence shown in Figure 5 was first edge detected using the Canny operator [3] to create a binary edge image, the input of which was fed directly into the CDHT. The Shape Sequence Template used to form the basis for the CDHT was the same as that used during the synthetic noise performance tests, except that the direction of motion was left-to-right, rather than right-to-left – a transformation requiring only a simple flipping of the Shape Sequence Template and a sign inversion of the DC component. The results of this test is shown in Figure 6, with the reconstructed Shape Sequence Template, scaled and offset by the parameters found using the CDHT, overlaid on top of the original image sequence.



Frame 1 Frame 7 Frame 13

Frame 19 Frame 25 Frame 31

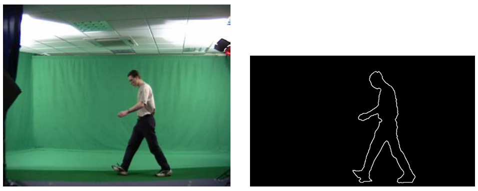
Figure 6: Results of the real-world data test

As can be seen from the results of the CDHT on real-world data, the target object is detected and tracked correctly, with the CDHT giving the correct spatial and temporal locations of the human walking in the image sequence. Only slight errors can be seen in the tracking of the subject and these can be put down to dissimilarities in the motion model of the Shape Sequence Template used by the CDHT and the actual motion of the human in the image sequence.

4 Human Gait Recognition

Access to the spatio-temporal frequency domain of the shape sequence provides a convenient method of analyzing fundamental structural properties of the shape sequence, and as such the spatio-temporal Fourier descriptors provide a good theoretical basis for discriminating between shape-sequences and therefore for classification and recognition of deformable objects.

To test the recognition capabilities of this technique, spatio-temporal Fourier descriptors were calculated for each subject in the Large Gait Database at the University of Southampton [14]. This database consists of 115 subjects and 1062 sequences (only right-to-left walking sequences were used for this test). In order to extract the boundary for each shape in an image sequence background extraction is first performed on each image (via chroma-keying), then each image is thresholded to produce a silhouette. The boundaries from each subject are then extracted by following the outer contour of each image, starting from the top of the head, to produce a complex boundary signal (see Figure 7), from which the spatio-temporal Fourier descriptors are calculated.



(a) Original Image (b) Extracted Boundary

Figure 7: Boundary Extraction

The descriptors produced using the method described above contain a large number of elements. In order to use these descriptors for classification purposes it is necessary to reduce the number of descriptors for each subject. This is necessary for two reasons, firstly to extract only the descriptors that would be useful for classification, and secondly to reduce the dimensionality of the feature space, thus ensuring feasible classification speeds.

The primary aim of feature selection in this case is to minimize intra-subject variance and maximize inter-subject variance, in order to increase the Correct Classification Rate (CCR). To achieve this one can use a variation of the Bhattacharyya distance metric to measure inter-class separation due to mean-difference with respect to the class covariances [5]. The separation between the two classes a and b, for a given feature, is given by

$$S_{a,b} = [m_a - m_b]\left[\frac{\sum_{\mathbf{a}} + \sum_{\mathbf{b}}}{2}\right]^{-1}[m_a - m_b]^T \qquad (14)$$

where m_a is the class mean and $\sum_a$ is the covariance matrix of class a, with equivalent terms for class b.

To gain a measure of a feature's ability to separate classes successfully a mean value of S was determined for each feature as

$$\bar{S} = \frac{1}{D^2}\sum_{a=1}^{D}\sum_{b=1}^{D} S_{a,b} \qquad (15)$$

where D is the number of descriptors available. $\bar{S}$ is then proportional to the class separability measure of the given feature, with larger values of $\bar{S}$ implying good class separability.

Classification testing used twenty features selected using the method described above as having the highest class separability values.

Classification was performed using a K-nearest neighbour classifier and cross-validated with the leave-one-out rule. This classifier assigns a test subject to be the same class as that of the modal class of the k^{th} nearest neighbouring subjects to it. If no modal class is found, then the test subject is assigned to the class of the nearest neighbouring subject. The distance between classes is measured by the Euclidean distance, ED, given by

$$ED = \sqrt{\sum_{n=0}^{N-1}(\mathbf{x_n} - \mathbf{y_n})^{\mathbf{2}}} \qquad (16)$$

where N is the dimensionality of the feature set, and $\mathbf{x_n}$ and $\mathbf{y_n}$ are the values of the n^{th} feature of the samples $\mathbf{x}$ and $\mathbf{y}$ respectively.

The classification results for $k = 1$ and $k = 3$ are shown in Table 1.

Table 1: Results of k-nearest neighbour classification

Database	k	CCR(%)
SOTON	1	84.5
SOTON	3	86.2

As can be seen, the selected spatio-temporal Fourier descriptors show a good ability at being able to discriminate between human subjects. These results compare favourably with other studies using the same database, with results of 82.9% and 71.2% for k=1 and k=3 respectively being reported by [4]. Hayfron-Acquah[8] reported recognition rates on the same database using a technique based around spatio-temporal symmetry of 92.8% for k=1 and 86.0% for k=3, thus showing an improved recognition rate over the technique discussed here for k=1, but not for k=3, suggesting that this technique produces features that demonstrate better clustering in the feature space.

The performance and robustness of using Fourier descriptors for the recognition of deformable objects was evaluated at varying resolutions, to simulate the effects of distance. As mentioned in Section 1.2, human gait as a biometric has the unique advantage of being useful at varying distances. To test the performance of spatio-temporal Fourier descriptors the effects of distance are simulated and performance testing is performed. The effect of distance is simulated by decreasing the spatial resolution of each image in an image sequence. The original images, which were at a resolution of 690 x 400, were scaled so that their heights were 128, 64, and 32 respectively. The results of classification at these resolutions, shown in Table 2, show that the image resolution can be relatively small without a great loss of resolution.

The fact that such a loss of spatial resolution results in only a small drop in recognition rate can be accounted for by the facts that the majority of the information in a given Fourier descriptor is contained in the low-level descriptors and that for a spatial boundary of length N we can obtain up to $\frac{N}{2}$ descriptors. Therefore, as long as we have a sufficiently large value of N, an adequate number of descriptors can be obtained for recognition. This requirement is more than fulfilled, even at the low resolutions used here.

Table 2: Results of k-nearest neighbour classification (k=3) for varying resolutions

Image Height	CCR(%)
400	86.2
128	84.5
64	85.7
32	82.4

5 Conclusions

The aim of this research was two-fold. In the first instance, the aim was to combine the powerful descriptive power of spatio-temporal Fourier descriptors with the robustness of the Hough transform, resulting in a new algorithm for the detection and extraction of deformable moving objects. The second aim was to test the capability of spatio-temporal Fourier descriptors to describe deforming moving shapes and to test their discriminatory power, in this case by applying them to the field of automatic gait recognition. In summary, spatio-temporal descriptors provide a new multiscale approach to describing temporally deforming shapes with the power to describe shape and deformation in a generalized way, suitable for shape extraction, and also in a more detailed way, suitable for shape discrimination.

Acknowledgments

We gratefully acknowledge partial support by the European Research Office of the US Army under Contract No. N68171-01-C-9002.

References

[1] J. K. Aggarwal and Q. Cai. Human Motion Analysis: A Review. *CVIU*, 73(3):428–440, March 1999.

[2] C. BenAbdelkader, R. Cutler, and L. Davis. Motion-based recognition of people in eigengait space. In *Proc. 5th FGR 2002*, pages 254–259, 2002.

[3] J. Canny. A Computational Approach to Edge Detection. *IEEE Trans. PAMI*, 8(6):679–698, 1996.

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

[5] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Morgan Kaufmann, 2 edition, 1990.

[6] D. M. Gavrila. The Visual Analysis of Human Movement: A Survey. *CVIU*, 73(1):82–98, January 1999.

[7] M.G. Grant, M.S. Nixon, and P.H. Lewis. Extracting moving shapes by evidence gathering. *Pattern Recognition*, 35(5):1099–1114, May 2002.

[8] J.B. Hayfron-Acquah, M.S. Nixon, and J.N. Carter. Automatic gait recognition by symmetry analysis. *Pattern Recognition Letters*, 24(13):2175–2183, September 2003.

[9] P. S. Huang, C. J. Harris, and M. S. Nixon. Recognising humans by gait via parametric canonical space. *Journal of Artificial Intelligence in Engineering*, 13(4):359–366, 1999.

[10] A. Johnson and A. Bobick. A multi-view method for gait recognition using static body parameters. In *Proc. 3rd AVBPA 2001*, pages 301–311, June 2001.

[11] A. Kale, A. N. Rajagopalan, N. Cuntoor, and V. Kruger. Gait-based recognition of humans using continuous hmms. In *FGR02*, pages 321–326, 2002.

[12] S. D. Mowbray and M. S. Nixon. Automatic Gait Recognition via Fourier Descriptors of Deformable Objects. In *Proc. 4th AVBPA 2003*, pages 566–573, 2003.

[13] J. M. Nash, J. N. Carter, and M. S. Nixon. Extracting moving articulated objects by evidence gathering. In *Proc. BMVC98*, volume 2, pages 609–618, 1998.

[14] J. D. Shutler, M. G. Grant, M. S. Nixon, and J. N. Carter. On a large sequence-based human gait database. *Proc. RASC02*, pages 66–71, 2002.

[15] J. D. Shutler and M. S. Nixon. Zernike Velocity Moments for Description and Recognition of Moving Shapes. In *Proc. BMVC 2001*, pages 705–714, 2001.