Extraction of Moving Articulated-Objects by Evidence Gathering

Jason M. Nash, John N. Carter and Mark S. Nixon
Department of Electronics and Computer Science
University of Southampton
Southampton SO17 1BJ, UK
[jmn95r|jnc|msn]@ecs.soton.ac.uk

Abstract

We present a new evidence gathering based approach, aimed to extract moving articulated objects from a temporal sequence of images. The new technique is designed to enable the automated determination of parameters pertaining to human gait, with a view to possible use as a biometric for recognition purposes. The articulated line feature extraction technique, uses a genetic algorithm (GA) based implementation of the Velocity Hough Transform (VHT). Using a GA to perform a heuristic search of the parameter space, rather than an exhaustive one, overcomes the problems of computation time and memory requirements associated with the original approach.

The new technique employs a parametric gait model consisting of a pair of articulated lines, jointed at the hip. Trials on real image sequences of pedestrians demonstrate that the approach is capable of locating and tracking a walking subject. Moreover, the technique is able to provide reasonable estimates of an individual's gait cycle period and hip rotation patterns, which are pertinent to recognition. However, current levels of accuracy are insufficient for these purposes. Nevertheless, the results demonstrate that the articulated line feature extraction technique has potential for use as an automated gait-data retrieval system.

1 Introduction

Human gait analysis is a subject into which there has been much research, with the majority of the work being conducted in the medical field. The aim of this research has been the identification of the basic components of human gait to enable the treatment of pathologically abnormal subjects, [9, 10]. More recently, an interest in gait as a biometric, "*a measure taken from a living person and used as a method of verification or recognition*", has developed, [3]. Studies have shown that if all gait movements are considered, gait is unique, [9, 10]. The uniqueness of human gait is attributed to the fact that the manner in which we walk is determined by our skeletal structure. Hence, as everyone has a unique structure and musculature, each individual has a characteristic way of walking, or gait signature, which can be employed as a means of identification.

Fundamental to all gait analysis problems however, is the acquisition of information pertaining to an individual's gait. In many gait-data collection systems, markers are attached to the subject to facilitate the measurement of the required information, [10]. However, such an approach is unsuitable for recognition systems as it requires contact with the subject. To overcome this difficulty, computer vision techniques have been adopted as a means of gait-data retrieval. The manner in which vision based methods are employed varies with the techniques developed, a brief overview of which is now presented.

1.1   Existing Techniques

Murase *et al.* propose a method for gait recognition which uses a parametric eigenspace representation of an individual's gait, [8]. The technique is a non-model based approach using spatio-temporal correlation, but appears sensitive to significant changes in illumination or background. Niyogi and Andelson, [12], use spatio-temporal patterns created by the pixels corresponding to a walking person over a sequence of images. Once again however, the technique is not directly related to body structure or the dynamics of a walking person.

A 3D model of a walking person is proposed by Marr and Nishihara, [7]. The model is a tree structured hierarchy of models corresponding to the identifiable parts of a human body. Hogg, [5], developed the *WALKER* model which represents a body as a collection of connected cylinders, and defines the temporal and spatial relationships between the cylinders for a set of postures which "*characterise the articulation of a walking person*". Rhor, [13], also uses Marr and Nishihara's model, combined with motion information gathered from medical studies. Again, instances of a model are individually matched to the first 10–15 frames of a candidate sequence and Kalman filtering techniques used to predict the model parameters in subsequent frames. As with Hogg's technique, Rhor's approach is capable of identifying pedestrians but not individuals.

It can be seen that computer vision based gait-data retrieval systems fall into two categories, either model based or non-model based. Non-model based approaches such as [8] and [12] do not consider the structure or dynamics of human gait directly, and use alternative representations of a walking person. Model based techniques, including [5] and [13] predominantly use the 3D model of Marr and Nishihara as a structural basis, coupled with data gained from medical studies detailing the dynamics of human gait. Paradoxically, current model based gait analysis systems tend only to be able to identify and track pedestrians, with the non-model based approaches being used to recognise individuals. Nevertheless, it has been suggested that an approach which is directly related to the mechanics of walking or body structure may offer improved recognition, [3].

To this end, one recognition measure based on Murray's work, [9], classifies an individual's gait by the observed rotation patterns of the hip and knee, [3]. Initial trials, with the required data being extracted manually, have shown that characterising gait in this manner has a significant potential for recognising individuals. Clearly, manual determination of the required gait-data is an undesirable process and difficulties can arise due to the self occlusion of the legs, resulting in missing data for those frames of the sequence. To overcome this problem, Cunado currently employs a curve fitting procedure to smooth the measured data and interpolate values for the missing segments. Accordingly, to achieve the desired aim of a fully automated system, a feature extraction technique is required to identify the position of the legs in each frame of an image sequence detailing

a walking person.

One such technique is the Velocity Hough Transform which was originally developed to locate moving objects from sequences of images, [11]. The VHT is a global evidence-gathering feature extraction technique which employs motion in the voting process of the Hough Transform (HT) as an inter-frame mapping. The inclusion of motion in the voting process provides several benefits, the first of which is that the motion exhibited by an object can be employed as a means of identification. The combination of motion with structural information pertaining to the feature of interest, bestows the VHT with improved feature extraction capabilities, compared with methods which consider structural information alone.

Inclusion of motion in the evidence gathering process also enables the integration of information present in each individual frame of an image sequence over the sequence as a whole. Hence, if information is missing in one frame, due to corruption by noise or the presence of an occluding object, the VHT is still able to accumulate a full description of the feature, as long as the information is present in another frame of the sequence. The VHT therefore, offers improved immunity to noise and resilience to the effects of partial occlusion, compared to feature extraction techniques which consider each frame of a sequence individually, [11].

Accordingly, being able to identify features with a particular structure which exhibit a certain type of motion from a sequence of images, the VHT is ideally suited to the problem of gait analysis. Furthermore, the VHT is a global technique which is not dependent on good initialisation. Moreover, the global integration of information that occurs over the whole image sequence, enables model parameters for frames where data is missing, e.g. due to the self occlusion of the legs, to be determined from the evidence that is gathered from the preceding and subsequent frames of the sequence. However, to enable the determination of the required gait-data the VHT must be reformulated to identify legs, which can be modelled as a pair of articulated lines jointed at the hip.

The following section introduces the basic concepts of the VHT and extends the theory to enable the extraction of pairs of articulated lines, i.e. legs. The paper continues by discussing certain implementation issues associated with the VHT, namely the computational time that is required and memory demands and attempts are made to address these issues through use of a genetic algorithm. A GA based approach is described, which offers equivalent feature extraction capabilities to the VHT, yet at a fraction of the computational requirements. Results of the new gait-data feature extraction technique are presented in section 3 and the conclusions which can be drawn from this work, together with possible avenues for future work, are discussed in section 4.

2   Velocity Hough Transform

The VHT was originally proposed in the context of extracting circles, moving with constant linear velocity, from a sequence of images. The polar parametric form is used to identify possible initial centre co-ordinates of a moving circle, given a range of radii and potential velocities to search over.

$$x = c_x + v_x t + r\cos\theta \qquad y = c_y + v_y t + r\sin\theta \tag{1}$$

Equation (1) is the polar parametric form of a circle moving with constant linear velocity, where c_x and c_y are the centre co–ordinates, r is the radius, x and y are the edge point

co–ordinates and θ varies from 0 to 360°, see figure 1. The velocity of the circle along the x and y axes, measured in pixels per frame, is denoted by v_x and v_y, respectively. t is the time reference of the current frame, relative to the initial frame of the sequence, where $t = 0$. By employing equation (1) in a similar manner to the standard HT, it is possible to determine the structural and motion parameters of any circle present within an image sequence, given that each frame is labelled with a time reference value, relative to the initial frame, and assuming that the circle exhibits constant velocity only.
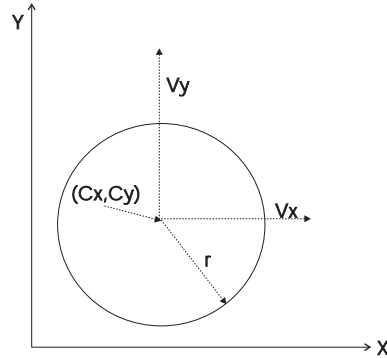
Figure 1: Parametric representation of a moving circle

Hence, the extraction of a circle moving with constant linear velocity requires a 5-D accumulator space to store votes for the parameters: c_x, c_y, r, v_x and v_y. In each frame, once an edge pixel is encountered, equation (1) is employed to identify the possible initial centre co–ordinates for a predefined range of radii and x and y velocities, given the time reference, t, of the current frame. The accumulator cells corresponding to each parameter combination, generated by the current edge pixel, are then incremented. Once all edge pixels in the current frame have been processed, the process repeats for the following frame of the sequence, voting in the same accumulator, until all frames of have been analysed. The result of the voting process is a peak in the accumulator space corresponding to the best estimate of the circle's parameters. As with the standard HT, multiple features lead to multiple peaks in the accumulator space, with each peak corresponding to the parameters of a particular feature.

However, to perform the desired task of locating the position of a pedestrian's legs in each frame of an image sequence, the VHT requires reformulation to extract pairs of articulated lines.

2.1   Articulated Line Extraction

Figure 2 shows a model of a pair of articulated lines, defined by the parameters: Ix, Iy, A, L and W. Ix and Iy are the x and y co-ordinates of the intersection point of the two central lines, respectively. A is the angle between the lines, L is the length of each line and W is the width. Accordingly, it is possible to approximate a pair of legs with such a model, where the intersection point of the two lines corresponds to the position of the hip. The dynamics of gait can then be modelled by varying the appropriate model parameters with time.
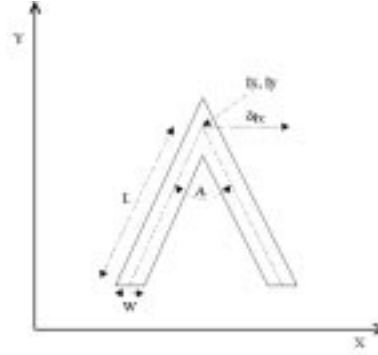
Figure 2: Articulated Lines

Human gait is periodic and medical research has shown that the rotation pattern of each thigh during a gait cycle is approximately sinusoidal in nature, [9]. Hence, variation of the model parameter A is also of a sinusoidal nature to parallel this behaviour.

$$A = B_A \sin(\omega_A \pi t + \phi_A) \tag{2}$$

B_A corresponds to the amplitude of the parameter variation with time t and ω_a and ϕ_A are the associated frequency and phase of the variation.

Similarly, a person's lateral motion can be modelled as a constant forward velocity coupled with a sinusoidal term, [9]; see equation (3).

$$Ix = V_{Ix} + B_{Ix} \sin(\omega_{Ix} \pi t + \phi_{Ix}) \tag{3}$$

Integrating this expression with respect to time leads to equation (4), which defines the absolute position of the model's intersection point, or hip, for a given time t.

$$Ix = Ix_{t=0} + V_{Ix}t + \frac{B_{Ix}}{\omega_{Ix}} \sin(\omega_{Ix} \pi t + \phi_{Ix}) - \frac{B_{Ix}}{\omega_{Ix}} \sin(\phi_{Ix}) \tag{4}$$

$Ix_{t=0}$ is the initial x co-ordinate of the intersection point and V_{Ix} is a constant velocity term. B_{Ix} is the amplitude of the sinusoidal variation and ω_{Ix} and ϕ_{Ix} are the associated frequency and phase. The final term of the expression is a constant of integration which ensures that there is zero displacement from $Ix_{t=0}$ at time $t = 0$. Motion in the y axis is currently not considered, i.e. a person is assumed to be walking on a flat surface parallel to the plane of the camera.

Clearly, this model is only an approximation, as consideration is not given to independent motion of the lower leg throughout the gait cycle, i.e. there are no knees. Nevertheless, the model is sufficiently accurate to determine the relative motion of the thighs, and corresponding hip rotation pattern, which is one of the major cues for recognition, [3]. However, the proposed model has 13 defining parameters which requires a 13 dimensional accumulator space in which to gather evidence. If implemented as previously described, the VHT becomes impractical for such a high dimensional parameter space, as the computation time and storage demands increase dramatically with the size of the problem. Attempts have been made to address this problem for the conventional HT which are directly applicable to the VHT.

One method is the use of extra data in the form of edge direction information to limit the number of accumulator cells addressed in the voting process, [6]. Moreover, edge direction information can be employed to decompose the parameter space of the HT incurring a saving in terms of both computational complexity and storage requirements, [1]. However, decomposition of the parameter space may lead to discrimination problems between features. Alternatively, the calculation of the HT's accumulator space can be re-ordered, [4], enabling subspaces of the Hough space to be examined in turn, resulting in a significant decrease in the amount of memory required to implement the technique. Although, this method offers no saving in terms of the overall computation time.

An alternative approach, is to employ a genetic algorithm. The VHT, along with all Hough based techniques, performs an exhaustive search of a given parameter space to determine the best estimate of the parameters of an instance of a particular feature present in an image. A GA however, performs a heuristic search of parameter space, incurring a significant saving in terms of the amount of memory and computational time required, without sacrificing resolution in the parameter space or the need for decomposition.

2.2   Genetic Implementation

A GA based VHT (GaVHT) employs a population of chromosomes, each of which represents an instance of the feature of interest, namely a model of the legs. The gait model has 13 defining parameters and hence each chromosome consists of a string of 13 genes, each of which is a binary representation of a particular model parameter.

The search for the optimal gait parameters is an iterative procedure where randomly initialised chromosomes are selected according to their fitness and combined, using the genetic operators *cross-over* and *mutation*, to produce a new generation of chromosomes. The fitness of a particular chromosome is determined by how well the feature template it represents matches the image data. As the VHT is concerned with extracting a moving pair of legs, the template is a multi-frame template which is matched across the whole image sequence. Hence, the fitness of a chromosome is determined by the number of pixels on the template it represents, which coincide with the image data. Strictly, this is actually the reverse of the VHT, where, like all Hough based techniques, image pixels are mapped to instances of a model, or template, and is the means by which the HT offers improved performance over conventional template matching. The genetic implementation however, maps instances of a model to the image data and hence performs conventional template matching. Nevertheless, the results are identical, [14].

The genetic operators *cross-over* and *mutation* are probabilistic operations which are performed on the binary bit strings of the individual chromosomes. *Cross-over* uses a pair of parent chromosomes and cuts them at a random position in their bit strings. The opposing sections of each parent's bit string are then recombined to form two children, e.g. consider the chromosomes 0100|1010 and 1011|0001. If the position between bits 4 and 5 is chosen as the cross-over point, joining opposing sections of the parents' bit strings results in the two new chromosomes 0100|0001 and 1011|1010. The *mutation* operator is then employed to randomly mutate the bit strings of the new chromosomes, complementing the affected bit. The probability of mutation is usually very low, of the order of $\frac{1}{1000}$, and is designed to ensure that some genetic variation propagates throughout subsequent generations.

As the parent chromosomes are selected in proportion to their fitness, only the fittest

chromosomes, or those whose gait parameters best match the image data, are used to create the next generation. Hence, after a number of generations the whole population is concentrated in the high performance section of the parameter space and eventually converges on the optimal solution, namely the best estimate of gait parameters for the subject present in the image sequence.

The next section presents the results of applying the GaVHT to the task of extracting the gait parameters of subject from a sequence of images.

3   Results

To establish the efficacy of using the GaVHT as the basis for a gait-data retrieval system, several experiments were conducted. These consisted of determining the position of the legs of a walking figure in each frame of a temporal image sequence. At this stage of development, the figure was imaged walking parallel to the plane of the camera against a plain background to facilitate the extraction process. Each frame of the gait sequence was digitised and then edge detected using an implementation of the Canny edge detector, [2]. The edge detected frames were then assigned a time index relative to the initial frame of the sequence and used as the input data for the GaVHT.

The extraction process begins by defining a search range and resolution for each of the gait model parameters, along with the size of the chromosome population. The first generation of chromosomes of the GaVHT are randomly initialised and then evolved (selected according to their relative fitness and combined probabilistically using *cross-over* and *mutation*) over a number of iterations. The evolutionary process is halted if the population satisfies a convergence criterion, i.e. the average fitness of the whole population is within a specified percentage of the maximum fitness, or a maximum number of generations is reached. The parameters corresponding to the fittest chromosome of the final generation are then taken to be the best estimate of the gait parameters for the subject in the sequence under analysis.

Figures 3 and 4 show the results of applying the GaVHT to sequences of two different individuals walking. It is clearly apparent, that in both cases the GaVHT has correctly identified the initial position of the subject and successfully tracked the figure throughout the sequence. Furthermore, the angular variation associated with the relative motion of the thighs has been well determined. This fact is demonstrated by the extracted position of the thighs in each frame of the sequence, although in some instances the influence of the lower leg has led to an error in the estimated position.

The absence of knees in the gait model is also a source of error, which is most evident when the subject's legs are crossing and one leg is bent. At this point, the results of the GaVHT favour the supporting leg, being the straighter of the two. The lack of knees, is also a contributing factor in the high estimate of the position of the intersection point, or hip. Nevertheless, despite the error in the positional estimate of the bent leg, the technique does correctly identify those frames of the sequence where the legs cross and thus successfully determines the period of the individual's gait cycle.

Figure 5 shows the extracted parameter variations for the relative thigh angle, *A*, and *Ix* displacement. Despite the approximate nature of the gait model, distinct differences are noticeable for the two subjects, suggesting that the technique has potential for recognition purposes.
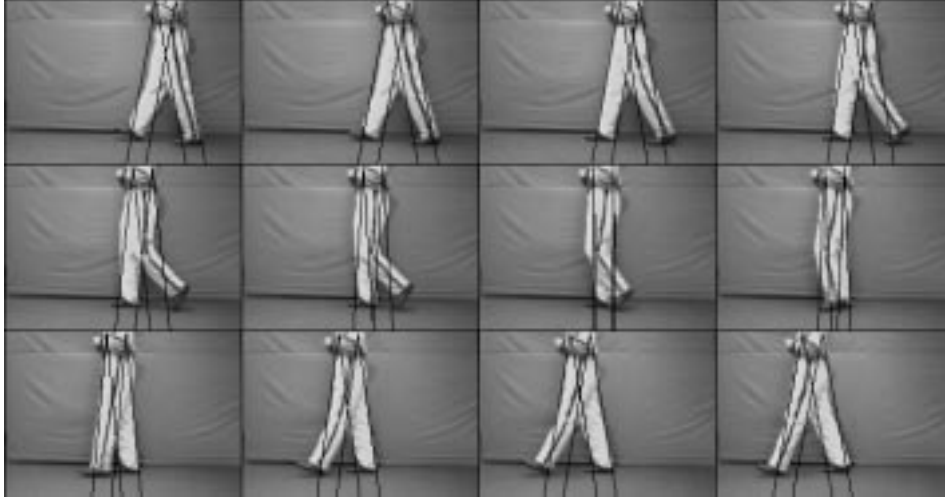
Figure 3: Extracted thigh data for subject 1. Sequences run from top left to bottom right.

To summarise, these results demonstrate that the GaVHT is capable of locating and tracking an individual throughout a sequence of images. Moreover, the technique can provide a good estimate of the period of the subject's gait cycle, in addition to an estimate of the position of each thigh, which is reasonably accurate for the majority of frames.

4   Conclusions

This paper has addressed the task of identifying the gait-data required to enable the recognition of individuals, through the use of computer vision based techniques. To this end, a GA based implementation of a global evidence-gathering dynamic feature extraction technique, the VHT, has been proposed. The GaVHT enables the problems of computation time and memory requirements, associated with all Hough based techniques, to be overcome, whilst retaining the fundamental traits of the VHT, namely improved feature extraction, immunity to noise and resilience to the effects of partial occlusion. In terms of performance, the GaVHT required approximately two hours on a 75MHz pentium to process the sequence in figure 3, whereas the results of the conventional VHT presented in [11], demanded of the order of days to compute for 7D parameter space applied to a ten frame sequence.

In conjunction with the GA based implementation of the VHT, an articulated line gait model has been developed to enable the extraction of the relative thigh positions of an individual throughout the gait cycle. The thigh positions can then be used to determine the hip rotation patterns of an individual, which are one of main cues used for gait recognition. The efficacy of the model has been assessed through experiments on sequences depicting a person walking. Results of these trials, demonstrate that the GaVHT is able to successfully locate and track an individual throughout a sequence of images. Moreover, the technique is able to identify the period of an individual's gait cycle and for the majority of the cycle, provide a good estimate of the relative thigh positions. However, for parts
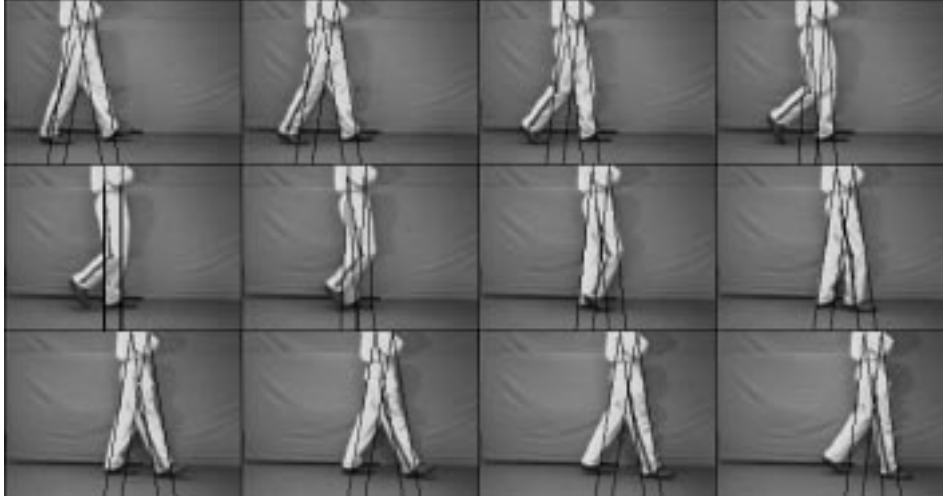
Figure 4: Extracted thigh data for subject 2. Sequences run from top left to bottom right.

of the gait cycle, most noticeably when the legs are crossing, the thigh position estimates can be erroneous, due to the influence of the lower leg and the absence of knees in the gait model. Nevertheless, identifiable differences in the extracted gait parameters for two subjects suggest a potential for recognition.

To conclude, the GaVHT gait-data extraction technique is able to accurately locate and track an individual throughout a temporal sequence of images. Furthermore, the algorithm is able to identify certain components of the gait cycle which are pertinent to the task of recognition, i.e. gait cycle period and hip rotation pattern, with a reasonable degree of accuracy. However, current levels of accuracy are insufficient for the extracted data to be employed as a means of recognising individuals. Hence, improvements to the gait model are required to attain the necessary level of confidence in the extracted information.

Therefore, the next progression in the development of a GaVHT based gait-data retrieval system is the inclusion of knees in the articulated line gait model. The addition of knees, will not only aid the extraction of the thigh data, but will also provide information relating to the knee rotation patterns, which is another the cue used for recognition.

References

[1] A. S. Aguado, M. E. Montiel, and M. S. Nixon. On using directional information for parameter space decomposition in ellipse detection. *Pattern Recognition*, 29(3):369–381, 1996.

[2] J. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[3] D. Cunado, M. S. Nixon, and J. N. Carter. Using gait as a biometric, via phase-weighted magnitude spectra. In *Lecture Notes in Computer Science Proc. AVBPA'97*, volume 1206, pages 95–102, 1997.

[4] G. Gerig and F. Klein. Fast contour identification through efficient Hough transform and simplified interpretation strategy. In *8th Int. Joint Conf. Pattern Recog.*, pages 498–500, 1986.
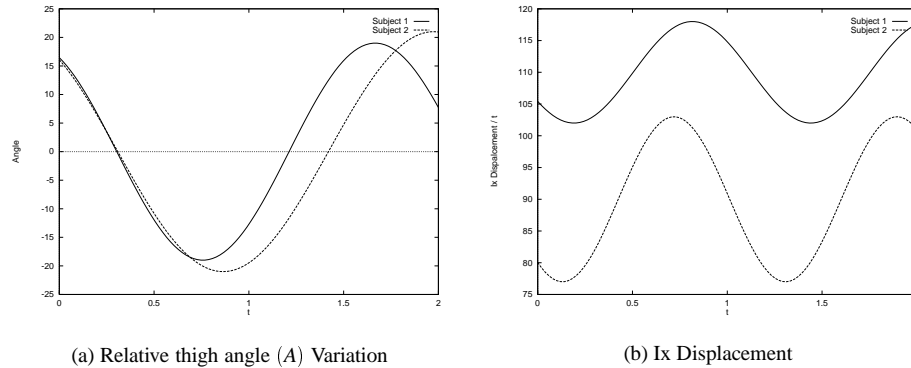
(a) Relative thigh angle (A) Variation

(b) Ix Displacement

Figure 5: Extracted gait parameters

[5] D. Hogg. Model-based vision – a program to see a walking person. *Image and Vision Computing*, 1(1):5–20, 1983.

[6] C. Kimme, D. Ballard, and J. Sklansky. Finding circles by an array of accumulators. *Communications of the ACM*, 18(12):120–122, 1975.

[7] D. Marr and H. K. Nishihara. Representation and recognition of the spatial organization of three-dimensional shapes. In *Proc. R. Soc. London B*, volume 200, pages 269–264, 1978.

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

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

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

[11] J. M. Nash, J. N. Carter, and M. S. Nixon. Dynamic feature extraction via the velocity Hough transform. *Pattern Recognition Letters*, 18:1035–1047, 1997.

[12] A. K. Niyogi and E. H. Adelson. Analyzing and recognizing walking figures in XYT. In *Proc. Conf. Comp. Vis. and Pattern Recog.*, pages 469–474, 1994.

[13] K. Rohr. Towards model-based recognition of human movements in image sequences. *Computer Vision, Graphics and Image Processing*, 59(1):94–115, 1994.

[14] J. Sklansky. On the Hough transform for curve detection. *IEEE Transactions on Computers*, 27:923–926, 1978.