Recognising Walkers Using Moving Light Displays

H. M. Lakany, A. Birbilis and G. M. Hayes
Artificial Intelligence Dept.
University of Edinburgh
5 Forrest Hill, Edinburgh, EH1 2QL
Scotland-U.K.
{hebaal,gmh}@aifh.ed.ac.uk

June 12, 1996

Abstract

In this paper, we present an algorithm to recognise walking people using moving light displays (MLDs). The algorithm adopted is based upon extracting the spatio-temporal trajectories of different joints of a subject. Features are then extracted using 2-dimensional fast Fourier transform and fed into a backpropagation neural network for classification.

1 Introduction

The fact that one can identify a friend at a great distance is an impressive human capability, as at great distances familiarity cues as facial features, hair style, clothing, etc. are practically obscure. Human gait motion has been of interest to researchers for decades. The development of photographic methods of recording a series of displacements of locomotion by the end of the nineteenth century encouraged researchers from different disciplines to study human motion and motion perception.

In the early experiments done by Marey [11], he observed an actor dressed in a black body stocking with white strips on his limbs. He studied the motion through observing the traces left on photographic plates as the actor walked laterally across the camera – an approach later known as *Moving Light Displays* (MLD). In [2], Braune & Fischer used a similar approach to study human motion, but they used light rods attached to the actor's limbs instead.

MLDs are image sequences containing only selected points of a 3-dimensional object in each frame. The method requires attaching a reflective tape or a light bulb to be attached to the joints/points of interest. If a reflective tape is used, then a strong light is focussed on the object. A video sequence is then recorded in which only patches of light can be seen by the viewer – see fig. 1.

Figure 1: a single frame of a walker having MLDs on joints

In 1973, Johansson [10], a psychophysist who was particularly interested in human motion perception, used MLDs in his experiments. He deduced that viewers cannot recognise any information from a single frame as in fig. 1, but can well recognise gaits when watching a sequence of such frames representing different activities such as walking, climbing stairs, dancing, etc. Cutting – Johansson's student – showed that using MLDs, one can recognise one's friends [5] and can determine the gender of a walker [6].

So, if humans can perceive gait with such a reduced spatio-temporal sequence, why not try to give a similar capability to machines?

Having a well defined structure that can be naturally decomposed into a hierarchy of parts, the human body was first modelled by Marr and Nishihara [12], a 3D model that was later used by Hogg [9]. Hogg proposed an approach that generated various poses of the given 3D model, projected them into the image and matched them with image edge data. His tracking procedure was based on examining the successive frames and averaging the size of clusters of points that are moving between frames with the assumption made about the position of the ground plane with respect to the camera.

Rashid [13] presented a computer system for MLDs which was able to track and cluster points belonging to independently moving objects. His tracking algorithm was based on the velocity of each point in a frame; he selected a correspondence that minimised the sum of the differences between the expected position and the actual position of a point in the following frame. Rashid filmed sequences of wire-frame figures moving so as to simulate a walking man, a dog and a moving toy-truck. The system was used successfully with one and two walking men, a man walking a dog and the moving toy-truck. Goddard [8] built a system that can recognise the gaits walking, jogging and running. He used the joint angles and angular velocities of the points of MLDs as input to his system. Bullpit & Allinson [3] describe a method that uses spatio-temporal neural networks to interpret motion in MLDs. In [7], the authors use object shape features and a spatio-temporal classification method based on a hidden Markov model to build an automated system for MLD recognition.

In this paper, we present an algorithm for recognising walking persons based on MLD sequences using 2D fast Fourier transform.

2 Experiments

The proposed approach is carried out in three phases: a *low level* phase where data is collected and the co-ordinates of the blobs are estimated, a *medium level* phase in which trajectories and features are extracted, and finally, a *high level* phase where recognition is performed.

2.1 Environment

The data collected for our experiments were taken according to the following conditions:

- Subjects are carrying no loads, i.e. only free normal walking is considered.
- Sequences are taken for subjects walking towards a stationary camera, i.e., the back and forth motion of the walker is not considered.
- Thirteen light bulbs are attached to each actor in the following distribution { 1 Head, 2 Shoulders, 2 Elbows, 2 Wrists, 2 Hips, 2 Knees, 2 Feet }
- Lights are always visible, i.e. no self occlusions are considered.

2.2 Data Collection and Preparation

Data collection starts with filming the subject then sampling the film to a sequence of frames of still grey-level images. For estimating the centre of blobs, images are thresholded such that if the intensity of a pixel is greater than the threshold, it is left unaltered, otherwise, the intensity value is replaced by the threshold value. Then the centres of the blobs are computed based on the weighted sum of intensities in the area that contains the blobs.

2.3 Scenario

Having estimating the co-ordinates of the MLD points in all frames, the following scenario is adopted.

1. Follow the points through the frames, i.e. find the locus of the spatio-temporal trajectory of each point.
2. Extract the features from the trajectories obtained.
3. Use the feature vectors for recognition.

2.4 Implementation

2.4.1 Tracking

Following the dots from one frame to the next and drawing the locus of the trajectories is the next stage.

Different tracking algorithms are mentioned in literature, varying in their level of complexity and accuracy – see for example [4] and [1]. Our algorithm uses a simple, yet reliable technique for tracking. The algorithm calculates the Manhattan distance of each pair of co-ordinates (of two successive frames) and then applies a labelling procedure to label the different co-ordinates in all frames – based on a priori knowledge about the human body structure and the position of the joints/lights. The tracking procedure developed in this work assumes smooth motion.

Figure 2 shows the locus of the trajectory of the head of a subject in both the $x-$ and $y-$directions.

2.4.2 Feature Extraction – 2D FFT

The features extracted are the components of the 2D FFT of each trajectory of each point – see fig.3. Components are then averaged by dividing their amplitudes by the number of samples.
Tracking and feature extraction are considered the medium level of the algorithm.

2.4.3 Dimensionality Reduction

Although it might be ideal to consider all components of the Fourier transform, it is certainly not practical, e.g. for a 3-cycle sequence, ~ 46 frames are captured, i.e. 23 feature components for each point, 23 x 13 points = 299 features : which is a huge number of input nodes for a neural network for such a simple classification task. Some criteria must be considered to reduce the dimensionality of the features.

Principal component analysis applied to examine the relevance of the Fourier components showed that higher order harmonics are of lower relevance. Hence, only

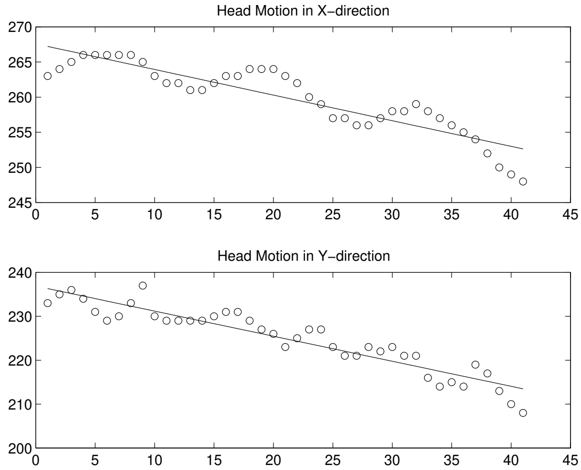

Figure 2: spatio-temporal trajectories of the head of a subject

the fundamental and the next few harmonics are considered as features.

Another kind of reduction can also be considered. It is based upon the nature of human walking and the coordination of the movement of the joints. Statistically speaking, one should estimate the value of the correlation coefficients between the different points. Points that have strong correlation with other points can be discarded. The correlation criteria requires large data sets to draw general rules for human movement behaviour.

2.4.4 Recognition

A neural network is trained using the fft components of the points for several sequence sets. Then, the network is used for recognition. This phase is the last phase of the algorithm.

A block diagram of the proposed algorithm is illustrated in fig. 4

3 Results and Discussion

The algorithm was applied to recognise two subjects – a male and a female. Based on principal component analysis, for each trajectory, we chose the fundamental and the next three harmonics as input to the neural network. The architecture of the network was: 52 input nodes (4 components x 13 points), 12 hidden nodes and 2 output nodes. Backpropagation was used to train the network. The learning rate was set to 0.1. We had *eight* sequences for each subject.

Table 1 shows the recognition results obtained.

Based on the experiments done, we would like to point out the following:

- The success of the whole algorithm depends greatly upon the accuracy of estimating the co-ordinates which in turn depends upon the underlying hardware. Perfect freeze-frame grabbing is essential, such that the blobs would not be smeared and hence, result in wrong estimation of the actual positions

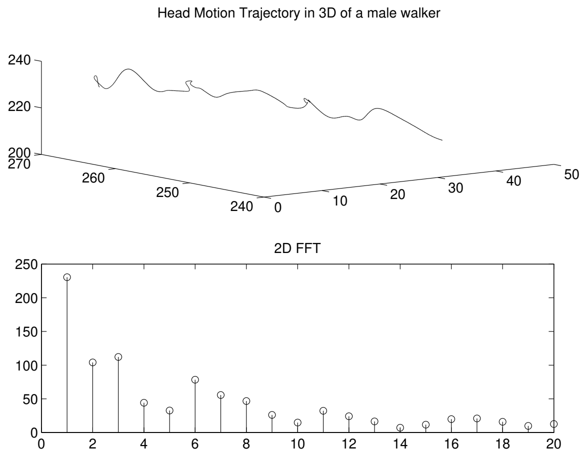

Figure 3: 2D FFT of head trajectory of a male subject

Table 1: Recognition Results for two subjects

	No. of Iterations	No. of Training sequences	No. of Testing sequences	Result Recognised/Total
1	26	14 (7+7)	2 (1+1)	2/2
2	26	12 (6+6)	4 (2+2)	4/4
3	24	10 (5+5)	6 (3+3)	6/6
4	21	8 (4+4)	8 (4+4)	7/8
5	19	2 (1+1)	14 (7+7)	9/14

of the MLDs. This consequently would affect the locus of the spatio-temporal trajectories.

- Determining the number of cycles (i.e strides) in a sequence is not an easy process. Human intervention was needed. Yet, for accuracy, this process should be automated.
- In our experiments with only two subjects, we did not reduce the dimensionality of the feature vector based on correlation, as it is difficult to deduce general rules from only two subjects.
- For more accurate results, more subjects and more sequences are needed.

4 Conclusion and Future Work

We have developed a method for recognising people from their walk based on features extracted from the 2-D fast Fourier transform of the trajectories of their major joints.

In future work we aim to:

- Investigate the possibility of using connectionist models (e.g. radial basis functions) to compensate for low sampling rate.

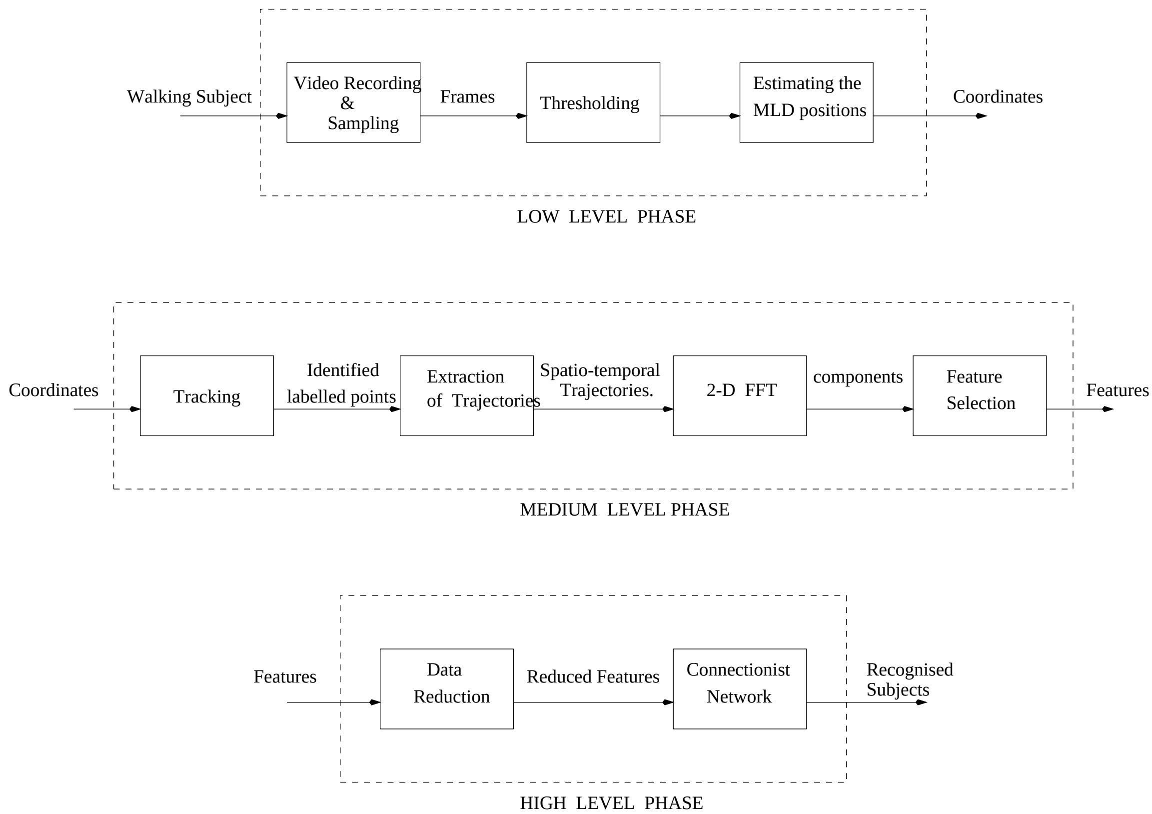

Figure 4: block diagram illustrating the main phases of the algorithm

- Automate the process of extracting a single cycle from a sequence.
- Develop a tracking algorithm which is able to account for occlusions.
- Take into consideration the walker's anthropometry and investigate the possibility of having individual walking signatures

References

[1] A.M Baumberg and D.C. Hogg. An efficient method for contour tracking using active shape models. Technical Report 94.11, University of Leeds - School of Computer Studies, Division of Artificial Intelliegence, April 1994.

[2] Wilhelm Braune and Otto Fischer. *The Human Gait.* Springer Verlag, 1987. Translated edition by P. Maquet and R. Furlong.

[3] A. J. Bullpit and B. Flinchbaugh. Motion perception and recognition using moving light displays. In *Second Int. Conf. on Artificial Neural Networks.*, IEE conference publication ; no.349, pages 91–94, November 1991.

[4] N.J. Byrne, A.M. Baumberg, and D.C. Hogg. Using shape and intensity to track non-rigid-objects. Technical Report 94.14, University of Leeds, School of Computer Studies, Division of Artificial Intelligence, 1994.

[5] J. E. Cutting and L.T. Kozlowski. Recognising friends by their walk: Gait perception without familiarity cues. *Bulletin psychonometric society. 9*, 1977.

[6] J. E. Cutting and L.T. Kozlowski. Recognising the sex of a walker from dynamic point-light displays. *Perception and psychophysics*, 1977.

[7] Kenneth H. Fielding and Dennis W. Ruck. Recognition of moving light displays using hidden markov models. *Pattern Recognition*, 28(9):1415–1421, 1995.

[8] Nigel H. Goddard. *The perception of articulated motion : recognizing moving light displays*. TR (Rochester) - 405, University of Rochester : Department of Computer Science, Rochester, NY, 1992.

[9] D. Hogg. Model-based vision: a program to see a walking person. *Image and Vision Computing*, 1(1):5–19, 1983.

[10] Gunnar Johansson. Visual motion perception. *Scientific American*, pages 76–88, November 1976.

[11] E.-J. Marey. *Movement*. William Heineman, London, 1895. Reprinted 1972.

[12] D. Marr and H.K. Nishihara. Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. Royal Society London B.*, 200:269–294, 1978.

[13] Richard F. Rashid. Towards a system for the interpretation of moving light displays. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-2(6):574–581, November 1980.