Analysis of Human Motion using Snakes and Neural Networks

Ken Tabb, Neil Davey, Rod Adams & Stella George

e-mail {K.J.Tabb, N.Davey, R.G.Adams, S.J.George}@herts.ac.uk
http://www.health.herts.ac.uk/ken/vision/
Tel: (+ 0044 / 0) 1707 286171
Fax: (+ 0044 / 0) 1707 284954

Department of Computer Science, University of Hertfordshire,
College Lane, Hatfield, Hertfordshire, UK. AL10 9AB

Abstract. A novel technique is described for analysing human movement in outdoor scenes. Following initial detection of the humans using active contour models, the contours are then re-represented as normalised axis crossover vectors. These vectors are then fed into a neural network which determines the typicality of a given human shape, allowing for a given human's motion deformation to be analysed. Experiments are described which investigate the success of the technique being presented.

Keywords: Human, Motion Analysis, Shape, Snake, Active Contour, Neural Network, Axis Crossover Vector

1   Introduction

This paper outlines a mechanism for analysing the motion and deformation of walking humans. A method based upon active contour models, snakes [1], and a neural network for categorisation is used [2, 3]. In previous papers we have discussed the classification of human shape in isolated static images taken from a motion sequence; in this study we discuss an individual's shape deformation during motion. The method discussed is part of a larger system designed to track moving pedestrians, a problem that has been the subject of much research [4, 5, 6]. We show that the periodic nature of human walking is clearly discernible from the deformation pattern, and that individual humans have a specific temporal pattern.

The paper is divided into 6 sections. Section 2 discusses the use of active contour models for detecting walking humans. Section 3 discusses the issues faced when using active contour data to train neural networks, and presents a solution in the form of the axis crossover vector. Section 4 summarises previous experimental findings using neural networks to identify human shapes, which leads into current work on human motion analysis using neural networks (section 5), along with experimental methods and results. Finally section 6 discusses the findings of this paper and details future work in this area.

2 Identifying and Tracking Moving Humans using Snakes

2.1 Snakes

In order to identify and track human outlines, the basic human shape must first be identified, and we use active contour models, 'snakes'. A snake is an energy minimising spline which can detect objects in an image and track non-occluded objects in a sequence of images. Snakes can be optimised for detecting and tracking particular classes of objects by customising their energy function so that detection of the desired characteristics, for example the curvature of an object's outline, or the presence of a particular colour, results in a reduction of the snake's overall energy. This reduction in energy has the effect of attracting the snake towards these desired features. Examples of a snake tracking a human can be seen in Figure 1. The active contour model used in these experiments was based on the Fast Snake model [7] as it allows for more autonomous object detection and tracking than the original model [1], although the techniques discussed in this paper are independent of the particular active contour model being used.
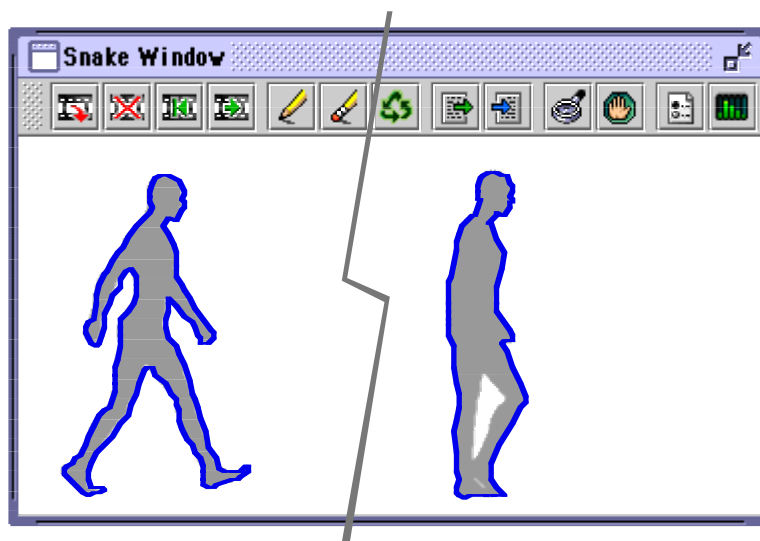

Figure 1: A target human detected and tracked in a sequence of frames, depicted here by the results of two discontiguous frames, using an active contour model. The target human has been dimmed in this figure to increase the snake's visibility.

Once the snake's energy function has been defined, the user clicks an initial polygonal contour around the target object in the first frame of a movie. The snake then minimises its contour's energy according to its energy function, causing it to lock onto features in the image which are defined as salient by the energy function. Once the snake has stopped moving it is copied, in its relaxed position, into the next frame of the movie, where it again minimises its energy in an attempt to lock

onto the target object in its new location in the image. This process continues until the end of the movie is reached. The success of the snake in detecting and tracking the target object is largely dependent on the relevance of the snake's energy function to the target class of object; the image quality following image preprocessing; the frame rate and speed of movement of the target object; the distance the contour's control points are from each other; and the level of occlusion of the target object. Once a snake loses focus from the target object, it rarely regains the successful detection and tracking of that target object.

A more detailed discussion of active contour models and their energy functions can be found in [8] and [9].

2.2   Tracking Walking Humans

In order to produce a clean and reasonably varied set of data, the 3D modelling and animation package 'Poser' [10] was used to simulate human movement. Snakes were relaxed around simulated humans in 30 different movies, all of which contained a single human walking from left to right. Each movie contained a different walk style and/or human build, providing a range of simulated human shapes and motions. All movies were 120 frames in duration, which allowed for at least 2 paces per person, despite individual differences in the length of stride and speed of movement from one human to another.

As described earlier, a snake is initialised around the human in the first frame of the movie. In subsequent frames, the snake is allowed to autonomously relax around the walking human, using its position in the previous frame as an initialisation in the next frame. In such a fashion the human is identified and tracked.
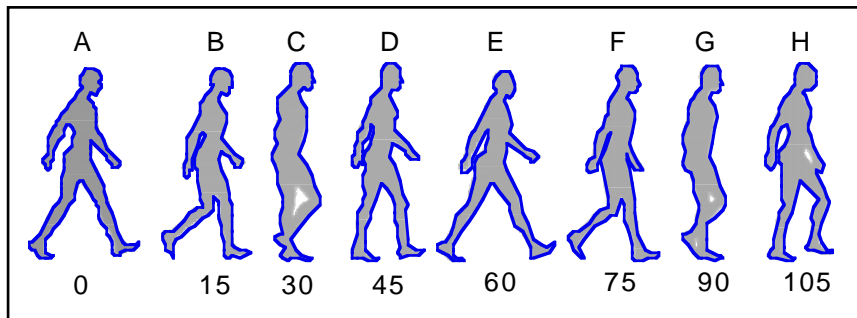

Figure 2: A target human detected and tracked through a sequence of 120 frames using an active contour. The frame numbers are shown below each frame; intermediate frames have been omitted. Above each frame is the pose identifier, referred to in the results sections of this paper. The target human has been dimmed in this figure to increase the snake's visibility.

Figure 2 shows the results of a snake tracking a sample human over 120 frames of video footage. No objective measure exists of an active contour's success

at tracking objects [8], but here the model was clearly successful as the humans were tracked in all of the 30 different instances. Speed is not a primary concern in this work, but it is worth noting that the relaxation times for the snakes in moving from one frame to another was often very fast.

As can be seen from Figure 2, the snake only obtains the outer edge of an object; 'holes' in between the human's legs do not form part of the contour. Despite the variation that humans adopt as they walk, their outer contour is identified in all cases.

3 Axis Crossover Vectors for Representing Human Shapes
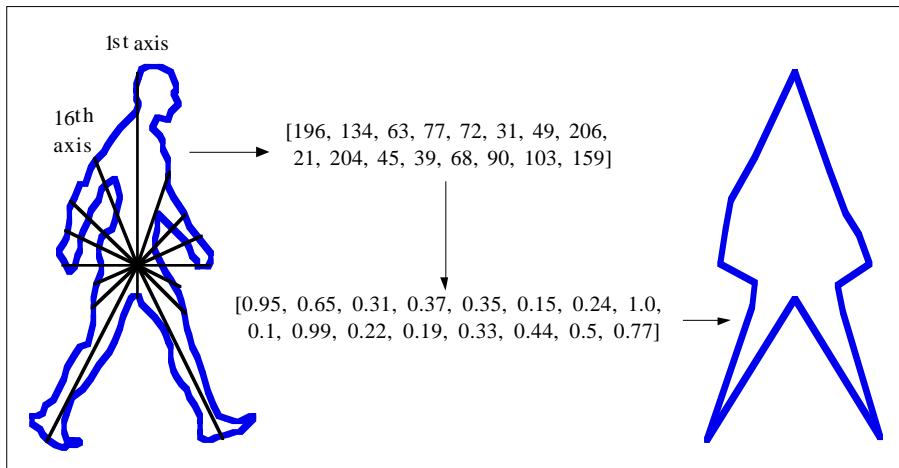
3.1 Representation

Figure 3: A human contour being re-represented as a 16-axis crossover vector. The snake's contour (left) has 16 axes projected from the contour's centre to its edges, giving 16 measurements (top vector). This vector is then normalised (bottom vector), to make it scale invariant, providing a compact representation for use in neural networks (visualised on right).

Once the target object has been detected and tracked in a movie, each frame's resultant contour can be stored as a vector of that contour's control points ((x,y) coordinates in the image), with vector length n where n is the number of control points along the snake's spline. However, this vector is not ideal for the purposes of shape analysis. By storing the absolute (x,y) locations of the control points, the active contour is location dependent; identically shaped contours in different parts of the image will have different vectors. The contour's vector is scale dependent too; differently sized contours which share the same shape will consist of different vectors. The contour is also rotation dependent; two identically shaped and sized contours will appear to have different vectors if they both consider different control

points to be the 'first', as it is this control point which appears first in the vector. Finally, the length of the vector is determined by the number of control points along the contour, therefore identically shaped and sized contours which contain different numbers of control points will result in vectors of different lengths.

All of these factors make the comparison and analysis of active contours difficult. For the purposes of being used in neural networks, the native active contour vector is additionally unsuitable because it contains pairs of data, the x and y coordinates of the control point, requiring the network to associate an x coordinate with the correct y coordinate before analysis can take place.
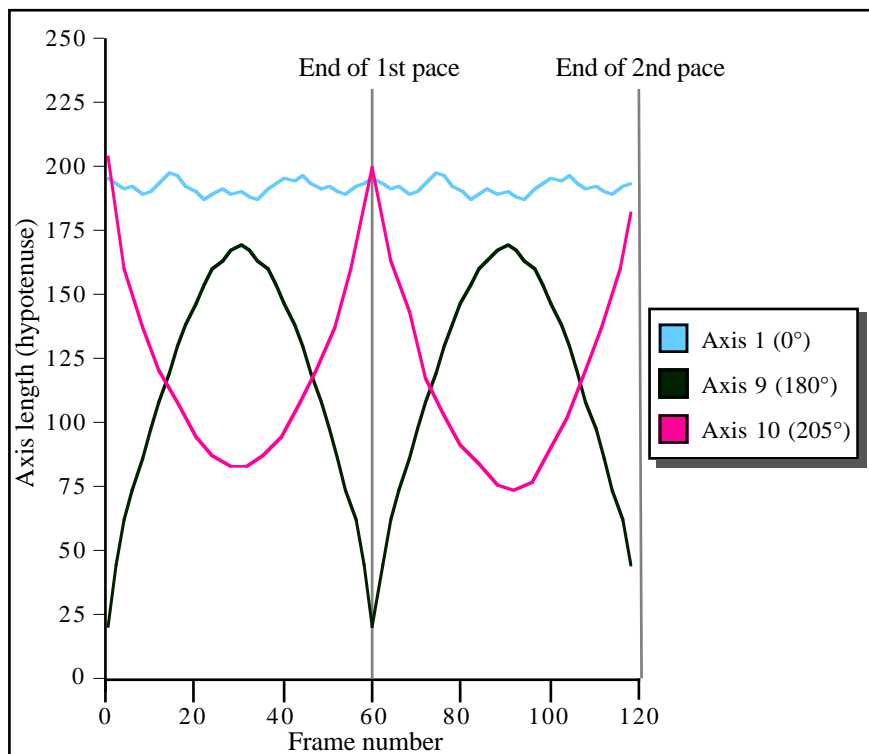

Figure 4: Axis length deformations over time, taken from a single walking human. The graph shows the differences in lengths of 3 of the 16 axes, at 0°, 180° and 205° from vertical, during two paces.

One possible solution to the problem of analysing active contours is the axis crossover representation, presented in [2] and shown in Figure 3. A number of axes are projected outwards from the contour's centre, with the distance along those axes from the centre to the furthest edge the axis meets being stored in a vector. This vector is then normalised by its largest element, making the vector scale, location and rotation invariant. Furthermore such axis crossover vectors are independent of the number of control points on the contour, as the crossover vector's length is

equal to the number of axes being projected, not to the number of control points on the snake. For more complicated situations the axis crossover representation offers more flexibility than discussed here; a more detailed description of its features can be found in [2].

3.2   Analysing Deformation

Having tracked the object with a snake, the relaxed contours are then converted into axis crossover vectors as described in Section 3.1. Figure 4 shows the deformation of 3 axes as a human walks along. Note that each axis' pattern of deformation is cyclical, closely repeating itself once per pace. Furthermore some axes are deformed more than others. The axis at 0°, which typically measures the distance from the contour's centre to the top of the head, shows little difference in movement during the walking motion. This reflects the fairly constant height that the head keeps from the centre of the body while walking. Conversely the axis at 180°, which typically measures the distance from the contour's centre to the crotch or feet (in the case where the legs are upright and the feet lie below the crotch), varies radically within each pace, reflecting the opening and closing of the human's legs. The length of the axis situated at 205° is has negative correlation to the 180° axis, where the opening and closing of the leg is again responsible for the axis growing and shrinking. Thus, when the leg is outstretched, the 180° axis is short whereas the 205° axis is long, and conversely when the human is standing up straight, the 180° axis is long and the 205° axis is short. Both the 180° and 205° axes extend and shrink only once per pace, coinciding with the stride being taken.

4   Identifying Human Shapes with Neural Networks

A range of neural network experiments have been performed to validate the axis crossover's ability to represent human contours in a scale-, location- and rotation-invariant manner [3]. Axis crossover vectors of different sizes were used, ranging from 4 axes to 24 axes, to identify the most appropriate axis crossover description for human shapes. The number of axes used has a direct relationship with the number of input units in the neural network: each axis' length is the input for one input unit. In the double output case one output unit was trained to represent human shapes, and the other non-human shapes. A range of hidden layers between 2 and $n-1$ units, where n is the number of input units in the network, were tested in order to identify the best generalisation ability.

Training data contained 150 human shapes, and 150 non-human shapes which consisted of 'outdoor furniture' shapes such as cars and streetlights. Active contours were relaxed around these shapes, and were then re-represented as axis crossover vectors, which could then be fed into the neural network.

Test data contained 10 unseen human shapes, and 10 unseen non-human shapes, again of outdoor furniture.

As is summarised in Figure 5, the experiments found that 16 axes offered the most suitable level of detail for encoding human contours using axis crossover

vectors. Consequently a 16 input unit network was chosen for all further experimentation. In addition, by using two units in the network's output layer, a confidence value can be obtained which allows a crude measure of how 'human' the network considers the given axis crossover vector to be. The confidence value is simply the difference of the two output units' values.
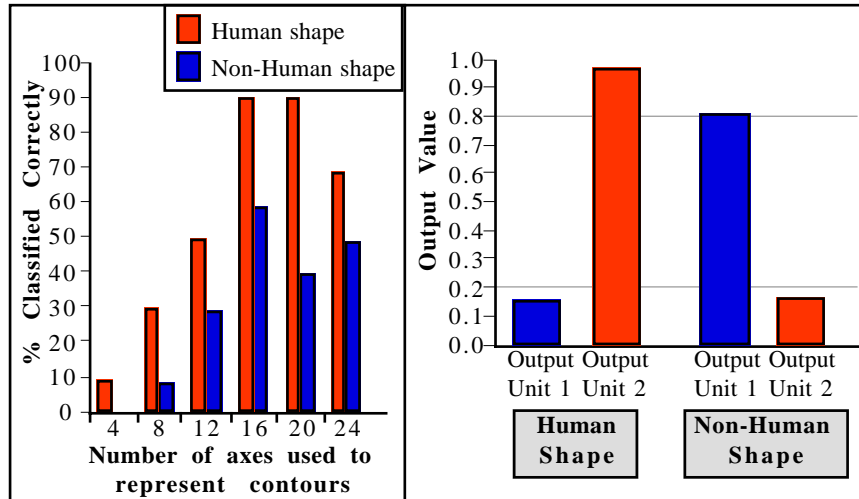

Figure 5: Results of neural networks classifying human and non-human axis crossover vectors. Left: Axis crossover vectors which used 16 axes were found to give the best results. Right: The mean confidence value when presented with human shapes (left pair of columns) and non-human shapes (right pair of columns).

5   Human Motion Analysis using Neural Networks

As described in the previous section, a trained neural network, when presented with a human shape represented as an axis crossover vector, can produce a scalar value measuring the confidence of the network that the shape is human. The network used in section 4 was tested with 30 different simulated human motion sequences, each varying in weight and / or gait.

The motion patterns are periodic, repeating here after the 60th frame, or at the start of the second pace. This can be seen in Figure 6, where two consecutive paces are superimposed for each of the 3 humans. The frequency of a human's motion pattern is attuned to the frequency of their walking, so that the speed of walking can be identified from the confidence value/time graph.

A human's motion pattern forms a signature specific to that human. Whilst the motion pattern for a human is not identical from one pace to another, in the same way that poses A and E in Figure 2 are not identical, it can be seen from Figure 7 that motion patterns for a given human are more similar to each other than to other humans' motion patterns.
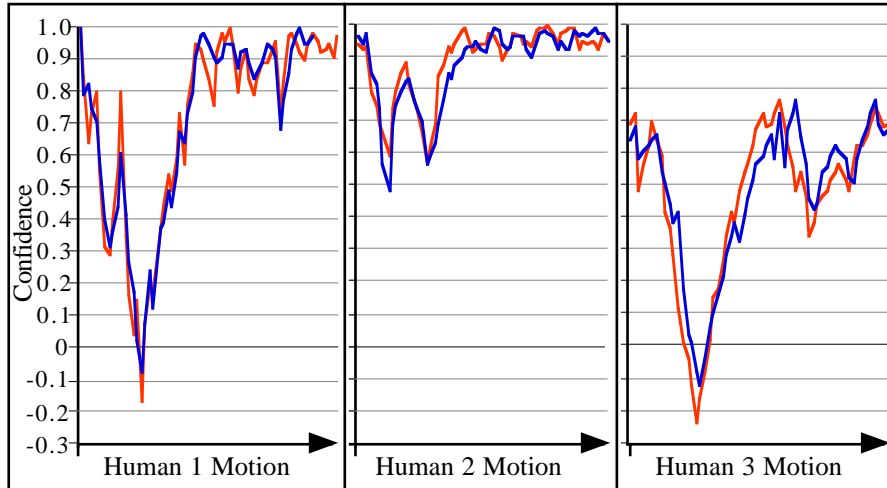
Figure 6: Analysis of human motion signatures. Each graph shows the neural network's confidence values over time for a given simulated human, with two consecutive paces for a corresponding simulated human being superimposed over one another. The graph on the left shows the signature for a simulated human with average overall confidence, whilst the middle and right graphs show the signatures for the simulated humans with highest and lowest overall confidence.
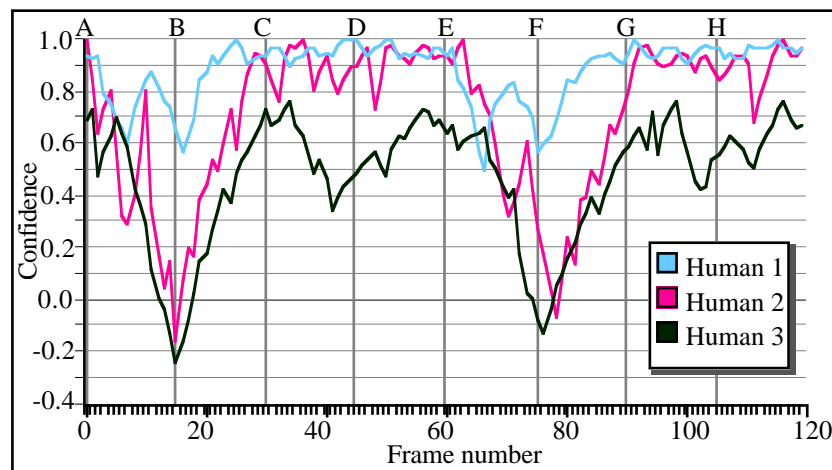

Figure 7: Neural network confidence value plotted against time for three simulated humans. 120 F rames corresponds to two paces. The pose identifiers labelled along the top of graph correspond to those in Figure 2.

As the human walks, this confidence value will change, reflecting the fact that the typicality of the shape of a walking human varies over time. This can be

clearly seen in Figure 7, where for some poses the network classes the humans as more non-human than human. See for example humans 2 and 3 in poses B and F. Moreover, there is considerable differences between humans so that, for example, human 1 is consistently classed as more human than humans 2 and 3 across all frames.

6 Discussion

In this paper we have shown how snakes can be used to identify human contours in an image, and then to track a moving simulated human. These contours can then be re-represented as fixed length, scale-, rotation- and location-invariant vectors. When the vector for a walking human is analysed over time, the deformation pattern of the human contour is observable , clearly showing a differential deformation pattern.

With a suitable set of images, a training set can be formed from the axis crossover vectors for a neural network classifier that identifies shapes as human or non-human, and it is the time dependent output of this network that can be used to analyse the motion pattern of a single walking human. As shown in Section 6, the output of the network is periodic, identifying single paces, so that speed of motion can be gauged. Moreover, individual differences between humans are apparent and these signatures could be used as identification tags. We are interested in extending this work to deal with the problems associated with occlusion. Our aim is to produce a model that will combine active contour models and neural networks and which will be able to track humans even during differing levels of occlusion, which will combine with previous studies on occluded human shapes [2]. This will involve prediction of the motion deformation pattern of a given human, which will in turn, involve the collection of a wider range of real human data, including footage of humans walking towards or away from the viewing point, and of real world noisy images.

References

1. Kass M., Witkin A. & Terzopoulos D. [1988]. *Snakes: active contour models*. In International Journal of Computer Vision (1988), pp. 321-331.

2. Tabb K., Davey N., George S. & Adams R. [1999]. *Detecting Partial Occlusion of Humans using Snakes and Neural Networks*. In Proceedings of the 5th International Conference on Engineering Applications of Neural Networks, 13-15 September 1999, Warsaw, pp. 34-39.

3. Tabb K., George S., Adams R. & Davey N. [1999]. *Human Shape Recognition from Snakes using Neural Networks*. In Proceedings of the 3rd International Conference on Computational Intelligence and Multimedia Applications, 23-26 September 1999, Delhi, pp. 292-296.

4. Sonka M., Hlavac V. & Boyle R. [1994]. Image processing, analysis and machine vision. Chapman & Hall.

5. Marr D. [1982]. Vision. W.H. Freeman & Company.

6. Baumberg A. M. & Hogg D. C. [1994]. An efficient method for contour tracking using active shape models. University of Leeds School of Computer Studies Research Report Series, Report 94.11.

7. Williams D.J. & Shah M. [1992]. *A fast algorithm for active contours and curvature estimation*. In CVGIP - Image Understanding 55, pp. 14-26.

8. Blake A. & Isard M. [1998]. Active Contours. Springer-Verlag.

9. Tabb K. & George S. [1998]. *Snakes and their influence on visual processing*. University of Hertfordshire Department of Computer Science Technical Report No. 309 Feb 1998.

10. MetaCreations [1999]. Poser 4 for Macintosh. www.metacreations.com