

# The Analysis of Animate Object Motion using Neural Networks and Snakes

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/>

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

**Abstract:** This paper presents a mechanism for analysing the deformable shape of an object as it moves across the visual field. An object's outline is detected using active contour models, and is then re-represented as shape, location and rotation invariant axis crossover vectors. These vectors are used as input for a feedforward backpropagation neural network, which provides a confidence value determining how 'human' the network considers the given shape to be. The network was trained using simulated human shapes as well as simulated non-human shapes, including dogs, horses and inanimate objects. The network was then tested on unseen objects of these classes, as well as on an unseen object class. Analysis of the network's confidence values for a given animated object identifies small, individual variations between different objects of the same class, and large variations between object classes. Confidence values for a given object are periodic and parallel the paces being taken by the object.

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

## 1 Introduction

This paper outlines a mechanism for identifying and classifying the shape deformation and motion pattern of walking humans and other animate objects. The method combines active contour models ('snakes') [Kass, Witkin & Terzopoulos 1988], with a categorisation neural network [Tabb *et al.* 1999a and 1999b]. A snake is used to detect the outline of an object in an image, having been manually initialised around the object by the user. Once the object's outline has been detected, the resultant contour is re-represented in a scale, location and rotation invariant vector. These vectors are then used to train a feedforward backpropagation neural network to distinguish human shapes from non-human shapes. Once trained, further unseen vectors are presented to the network to determine how accurate the network is at classifying human and non-human shapes. Snakes are then used to track animate objects in sequences of images, and the subsequent shape vectors are presented to the network to identify patterns in a given object's shape deformation. A discussion of these deformation patterns is given, along with possible uses of the technology.

The techniques presented are part of a larger system designed to track moving pedestrians, a problem that has been the subject of much research [Baumberg & Hogg 1994; Bowden, Mitchell & Sarhadi 1998; Galanta, Johnson & Hogg 1999]. 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.

## 2 Identifying and Representing Moving Objects

In order to analyse an animate object's deformable shape during motion it is first necessary to obtain the shape of that object in each of a series of images. Snakes are an established method (Blake & Isard

1998) for identifying and tracking moving objects. Snakes are energy minimising splines whose accuracy at detecting an object is dependent upon the how suitably defined the snake's energy function has been. A snake's energy can contain different types of shape criteria, such as the goal to stay circular, as well as different image criteria, such as having lower energy associated with, and thus being attracted to, edges or specific colours in the image. The combination of these criteria results in an energy function tailored to a specific task.

The user initialises a snake in an image or movie frame, and the snake relaxes around the target object until it finds a local or global minimum in the energy space. The snake, once relaxed in a minimum, can then be introduced into the next frame of a movie, and re-minimised, to track a moving object autonomously (Figure 1).

An active contour model based on Fast Snakes [Williams & Shah 1992] has been designed for the task, and provides a reasonably accurate means of identifying and tracking an object's shape during motion (Figure 2).

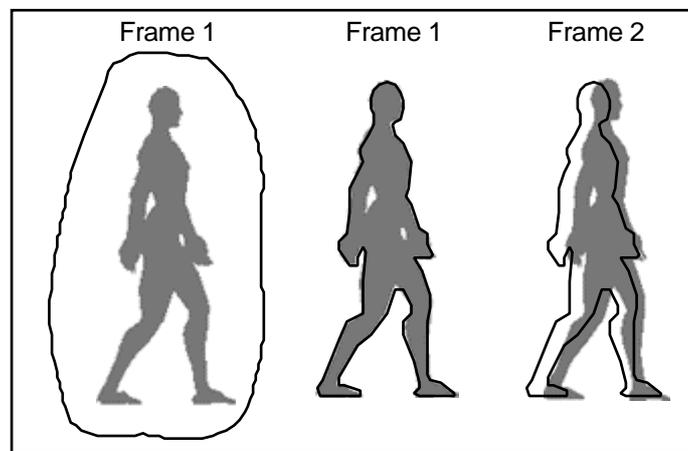


Figure 1: Detecting and tracking objects using an active contour. [Left] The user initialises a contour around the target human. [Middle] Minimising the snake's energy forces the snake to relax onto the human outline. [Right] Once relaxed, the snake is initialised in the next frame using its relaxed position as a starting point. Its energy is then minimised again to relax it onto the human's new position.

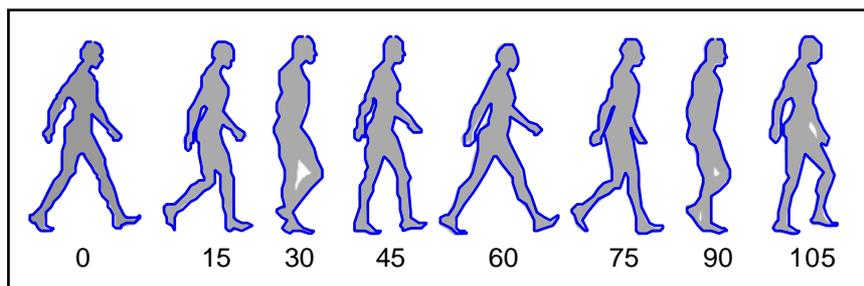


Figure 2: Human poses (grey) being tracked with an active contour (black). Numbers beneath each pose denote the movie frame number.

Due to an active contour's scale, location and rotation dependence, and additionally because it is stored as pairs of (x,y) coordinates, the contours are not suitable to use as input patterns for neural networks. Instead, an active contour can be translated into a scale, location and rotation independent axis crossover vector [presented in Tabb *et al.* 1999a]. A number of axes are projected outwards from the contour's centre (Figure 3), with the distance from the contour's centre to the furthest contour edge along each axis being stored in a vector. This vector is then normalised by its largest element, making the vector

scale, location and rotation invariant.

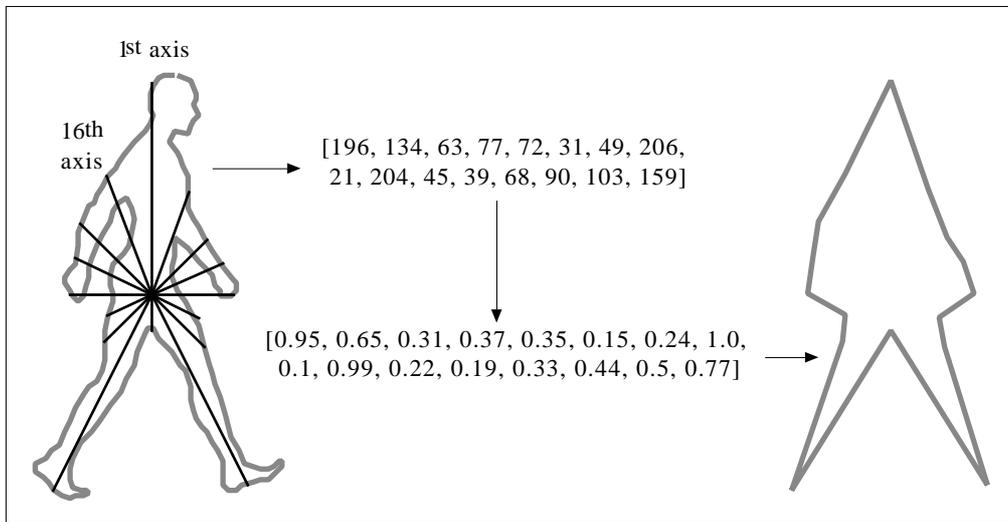


Figure 3: Converting an active contour into an axis crossover vector. Axes are projected at specific angles from the contour's centre to its edges. These axes' distances are then stored in a vector and normalised, making the vector scale, location and rotation invariant.

### 3 Identifying Human Shapes with Neural Networks

A range of experiments have previously been performed to validate the axis crossover's ability to represent contours sufficiently for a neural network to be able to distinguish human from non-human shapes [Tabb et al. 1999b]. This paper uses exclusively a network with 16 input units and 2 output units, where one output unit was trained to identify human shapes, and the other non-human shapes. All axis crossover vectors used contained 16 elements, where a given element maps onto a given input unit in the neural network (Figure 4).

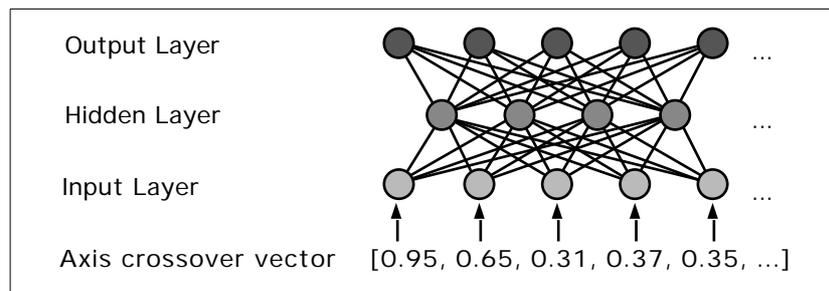


Figure 4: Axis crossover vectors as input patterns for neural networks. A given vector element maps onto a given input unit in the neural network, thus the size of input layer and axis crossover vector must equate.

In order to produce clean and reasonably varied sets of training and test patterns, simulated shapes of humans and non-humans were used. The 3D modelling and animation package Poser [Metacreations 1999] was used to generate all simulated human, equine, canine and velociraptor movement. The training set contained 400 simulated human and 400 simulated non-human shapes. The non-human shapes consisted of inanimate objects such as trees, cars and streetlights; and deformable animate objects, namely shapes of dogs and horses. A breakdown of the training set can be seen in Figure 5. All animate movement consisted of still images or movies of the object (human, dog, horse or velociraptor) walking from the left side of the image to the right. Each image or movie was different,

either in terms of the physical build of the object, for example height and weight of object, or in terms of its walking habit, for example one human might swing their arms more than another. Generating this data using Poser allowed for much more variation in shapes and motions than could be achieved in a reasonable time frame using real images.

Once trained, a scalar confidence value was obtained from the two output units which allows a crude measure of how ‘human’ the network considers a given vector to be. The confidence value is simply the difference of the two output unit values.

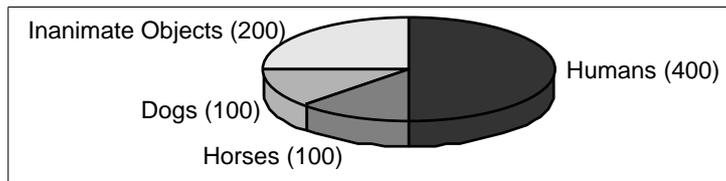


Figure 5: Training data set for the neural network.

Two experiments were performed in this study: a categorisation experiment to determine how accurate the network was at classifying static shapes as human or non-human, and a tracking experiment to analyse a given object’s deformable shape during motion.

In the training set, the animate objects were divided into bipedal humans and quadrupedal dogs and horses. In order to explore the nature of the induced classification, we introduced another bipedal class to act as ‘near-miss’ humans. The only other animated biped available in the software used was a velociraptor, a roughly human sized bipedal dinosaur.

### 3.1 Categorisation experiment

The categorisation experiment involved presenting the trained network with a test set of both human and non-human axis crossover vectors and measuring how successfully the network categorised them, based on the network’s confidence value for each shape. The test set for the categorisation experiment contained 100 unseen human shapes, and 100 unseen non-human shapes (see Figure 6a).

### 3.2 Tracking experiment

The tracking experiment involved tracking human and animate non-human objects using active contours, and for each object, converting the sequence of relaxed contours into axis crossover vectors, then presenting each vector to the neural network serially. A confidence value versus time graph can then be plotted, depicting how human the object was considered to be in each frame of its motion. The test set for the tracking experiment contained 30 unseen humans and 15 unseen non-human animate objects (see Figure 6b). Each object was tracked for 2 consecutive paces.

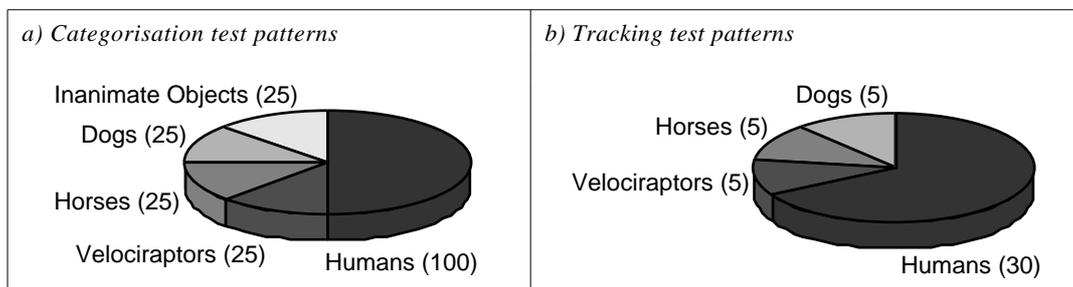


Figure 6: Test data set for the categorisation [left] and tracking [right] experiments. No inanimate object shapes were used in the tracking experiment. All shapes used as test patterns were previously unseen by the neural network.

## 4 Animate Object Motion Analysis using Neural Networks

### 4.1 Categorisation experiment

The results for the categorisation experiment on unseen data can be seen in Figure 7. As has been observed in previous experiments [Tabb *et al.* 1999a and 1999b], the network learns to classify unseen data very successfully. When the novel near-miss velociraptor class is presented, the results show that the network considers it to be more ‘human’ than ‘non-human’. Nevertheless it is not classified with as great a level of confidence as the genuine humans. In other words the network identifies the ‘near-miss’ class as a near-miss.

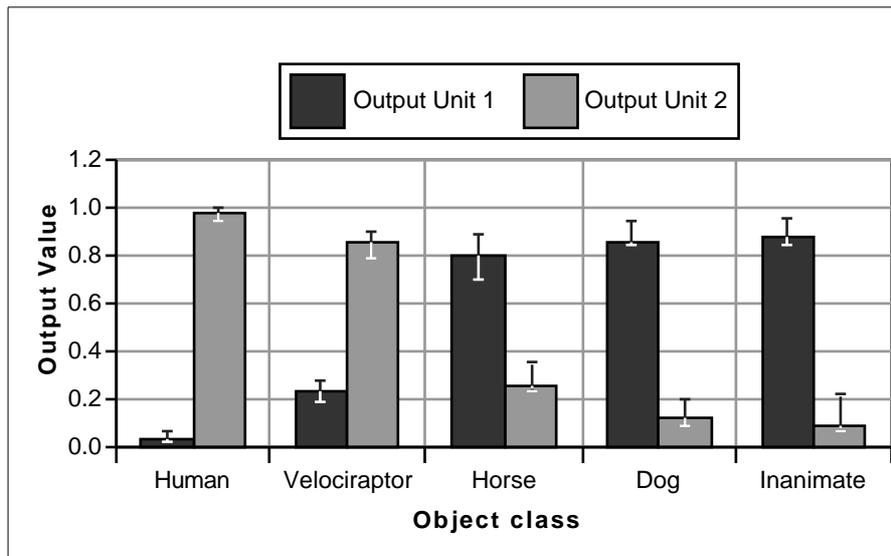


Figure 7: Analysis of neural network categorisation values for different types of unseen objects. The network’s confidence value for a given shape can be obtained by subtracting the first output unit’s value from the second. Results shown reflect the mean values given for 100 unseen human objects, 25 unseen velociraptors, 25 unseen horses, 25 unseen dogs and 25 unseen inanimate objects. Also shown are the standard deviations for each object class.

### 4.2 Tracking experiment

The results for the tracking experiment can be seen in Figure 8. Graphs are shown for 3 of the 30 humans in the test set, namely those which the network categorised with most, average, and least confidence. Graphs are also shown for the velociraptor, dog and horse categorised with average confidence for their respective object class. The two consecutive paces tracked for a given object have been superimposed on each object’s graph.

For each test object, the network’s confidence value varies with the object’s motion and is cyclical, with the frequency being once per pace. Furthermore, whilst a given object’s confidence pattern is not identical from one pace to another due to slight variations between paces, its confidence patterns from pace to pace were more similar to each other than to those of other objects in the same species, identifying individual differences apparent in objects’ movement.

These differences were even more marked when comparing objects from different species; Figure 9 shows the same objects’ confidence values being plotted against each other, with each object’s two consecutive paces being plotted end to end. It is evident that the neural network considers some species to be more human than others, as was the case in the categorisation experiment. Furthermore bands can be drawn across the graph, marking off each species from each other.

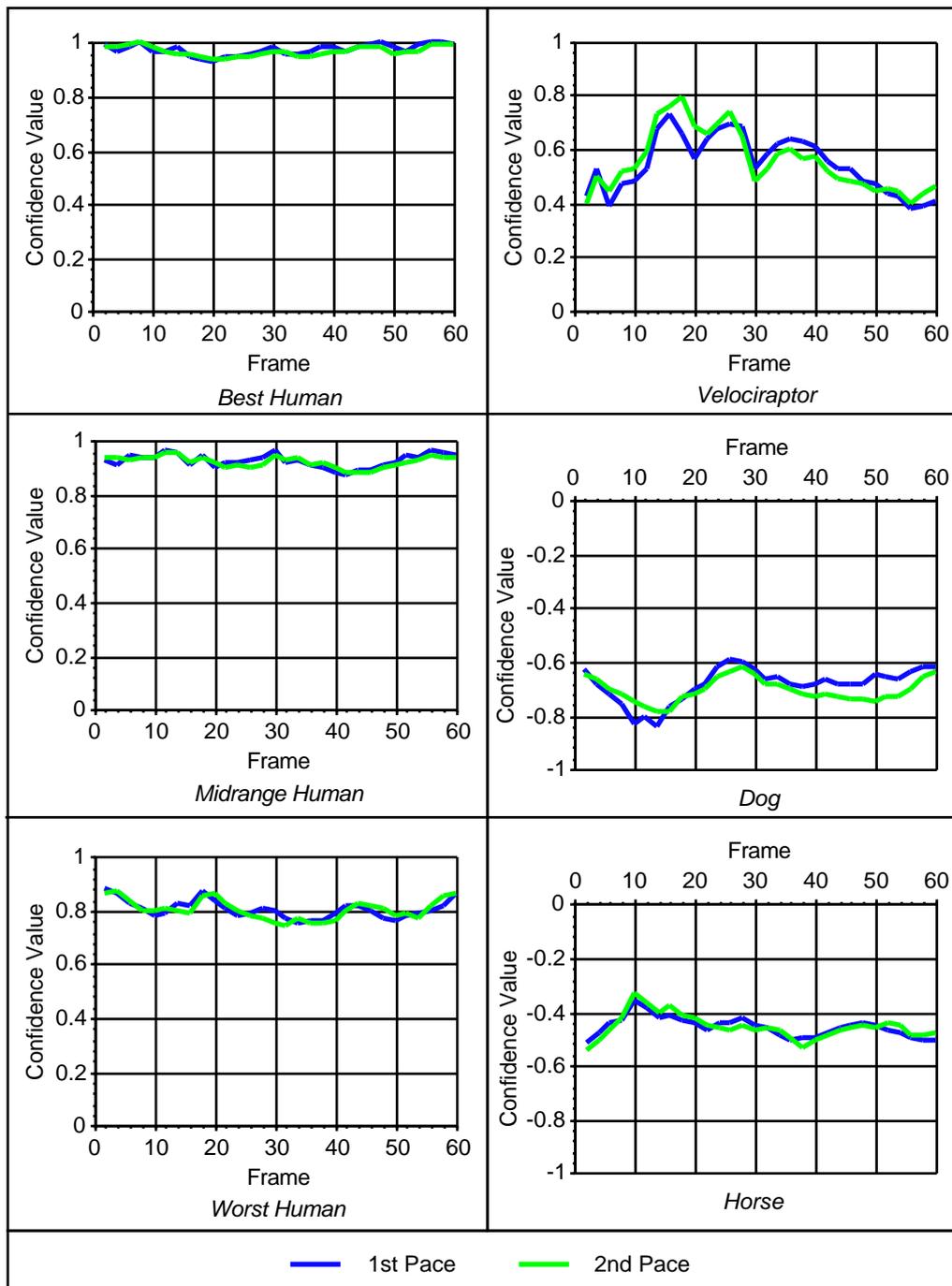


Figure 8: Motion analysis of animate objects. The three graphs on the left show best, typical, and worst human categorisation over time. The graphs on the right show the same information for a typical velociraptor, dog and horse. In each graph, 2 consecutive paces for the appropriate object have been plotted, showing more clearly each object's repetitive yet distinctive motion pattern.

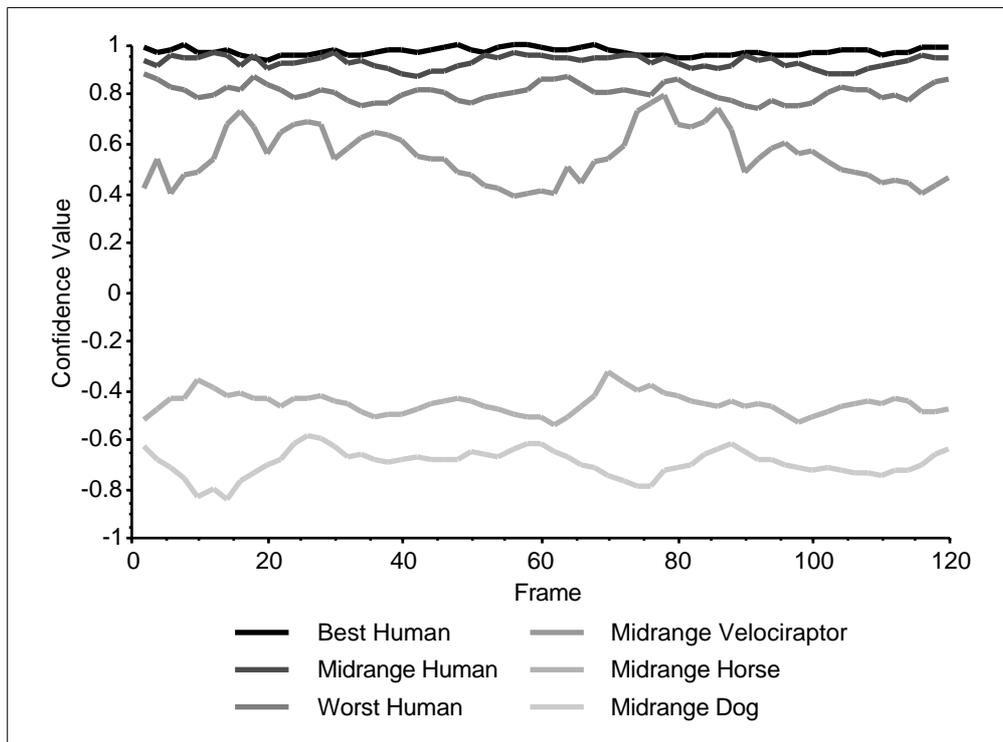


Figure 9: Comparison of neural network confidence values during animate object motion. Each object has been tracked for two paces, where the second pace starts at frame 61.

## 5 Discussion

In this paper we have shown how snakes can be used to identify an object in an image, and then to track a moving object. The resultant contours can be re-represented as scale, location and rotation invariant vectors. A neural network was trained using these vectors and was then able to successfully classify a range of unseen objects. The confidence value generated by the trained neural network for snapshots of a moving object form a periodic waveform individual to that object. A more detailed analysis of the network's confidence values for each given animated object class identifies small, individual variations between different objects of the same class, and large variations between object classes.

One possible use of the object's periodic waveform is to aid in identifying single paces of the object, so that the object's speed of motion can be gauged. Moreover, individual differences between objects are apparent and could be used as identification tags.

Other applications of the object's periodic waveform might be to identify the object's species, or to aid in predicting the object's imminent motion by 'fast-forwarding' along its waveform.

## References

- 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.
- Blake A. & Isard M. [1998]. Active Contours. Springer-Verlag.
- Bowden R., Mitchell T.A. & Sarhadi M. [1998]. *Reconstructing 3D Pose and Motion from a Single Camera View*. In Proceedings of BMVA 1998.

Galanta A., Johnson N. & Hogg D. [1999]. *Learning Behaviour Models of Human Activities*. In Proceedings of BMVA 1999.

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

Metacreations [1999]. Poser 4 for Macintosh. <http://www.metacreations.com/>

Tabb K., Davey N., George S. & Adams R. [1999a]. *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.

Tabb K., George S., Adams R. & Davey N. [1999b]. *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.

Tabb K., Davey N. & Adams R. [1999c]. *Analysis of Human Motion using Snakes and Neural Networks*. To be published in Proceedings of Articulated Motion & Deformable Objects 2000, Palma de Mallorca, September 9-10 2000. Submitted in December 1999.

Williams D.J. & Shah M. [1992]. *A Fast Algorithm for Active Contours and Curvature Estimation*. In CVGIP - Image Understanding 55, pp.14-26.