Statistical Models of Object Interaction

R. J. Morris, D. C. Hogg
School of Computer Studies, University of Leeds, England
rjm@scs.leeds.ac.uk, dch@scs.leeds.ac.uk

Abstract

We present a method for assessing the likelihood of a trajectory of an object through a scene consisting of a number of other objects. The closest points on the trajectory to the other objects are chosen as landmark points and at each landmark we calculate the probability of the interaction. Sequences of such probabilities are then analysed to give the likelihood of the whole trajectory.

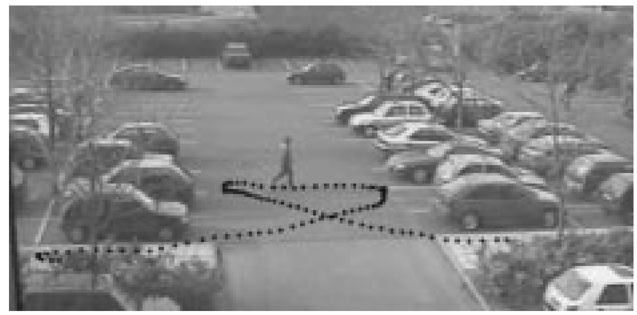

Figure 1. The car park scene and one of the trajectories through it

1. Introduction

The aim of our work is to devise general techniques for detecting atypical behaviours in the interaction of objects within a chosen domain. This paper describes a simple geometric method for characterising the relative movements of mobile and stationary objects within a probabilistic framework. Experimental results are presented for the domain of car park scenes in which people move to and from exits and parked vehicles. A probabilistic model of the interaction between people and cars is constructed automatically through observing long training sequences of video. Finally, this model is used to detect atypical behaviours that are plausibly significant within a surveillance context.

As a pre-requisite for the work undertaken, two model-based systems are used to track the motion of people and vehicles. The first [**?**, **?**, **?**] uses active shape models to track non-rigid objects, in our case people. The second [**?**] uses geometric 3D models to track rigid objects: cars. These systems have been integrated to handle mutual occlusion [**?**]. Both systems provide coordinates on the ground plane of the objects being tracked. It is this information that we use as our input data.

Several authors have devised systems to interpret the motion of objects through a scene. One approach is to construct a natural language description of the behaviour of the objects in the scene [**?**, **?**]. Probabilistic or Bayesian networks have been used to help construct such descriptions [**?**, **?**]. An alternative approach [**?**] is to model the probability density function of instantaneous movement and partial trajectories. All the above approaches rely on having a constrained scene where only a limited number of paths are taken.

We can view the task as one of compressing the data: we want to find ways of representing the data which preserves the useful information but is independent of the particular arrangement of objects. Once we have such a representation we can use standard statistical techniques to characterise the representations. Much of the data on a path contains little useful information, so we select certain landmark points which capture the interaction between objects. At these points we measure various quantities which describe the nature of the interaction. We then model the distribution of these quantities which enable us to calculate the probability of the interaction. This leaves us with a sequence of probabilities from which we can factor out the temporal information to give us a monotonically increasing sequence. Finally we use weighted sums of probabilities in a supervised learning framework to classify the types of behaviour.

The system has been tested on a set of data gathered from the car park shown in figure 1. This data consists of 54 trajectories collected during the morning rush hour, most of which involve some interaction with the vehicles. Five atypical trajectories have also been recorded. For this work the positions of the cars have been estimated by hand. For the purpose of our algorithm this is a reasonable approximation. In the final system we would use car positions automatically

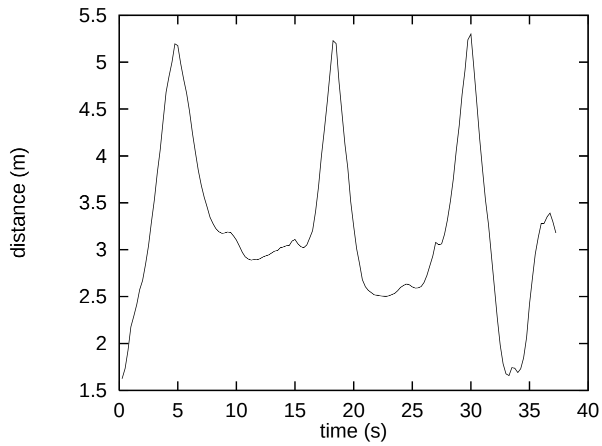

Figure 2. The distance to the currently nearest car

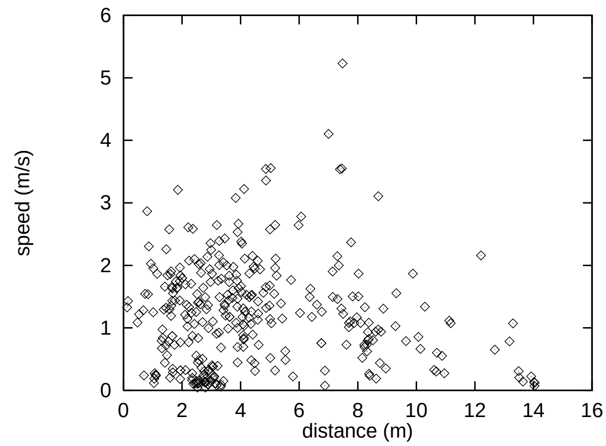

Figure 3. The distribution of speed and distance

generated by the vehicle tracker.

2. Choosing Landmarks

By choosing geometrically significant points or *landmarks* we significantly reduce the amount of information used to describe a curve. Landmark data has been widely used for statistical shape analysis [**?**] particularly in the medical field. In our situation we choose those points on a trajectory which are at a minimum distance from a specific object. In terms of the interaction between objects, these minima capture some of the most important information about a trajectory. In other applications we could use other geometric features of a curve such as vertices or inflections. For example Bookstein [**?**] has recently developed a novel statistical approach which uses bending energy of interpolating splines to specify landmarks.

Consider an object moving along a straight line with constant speed s past the origin where the minimum distance to the line is h, i.e. on the curve (h, st). The distance from the origin is given by

$$d = \sqrt{h^2 + s^2t^2} \approx h + \frac{s^2}{2h}t^2,$$

so it can be locally approximated by a parabola, for which it is easy to calculate the minima. This parabola becomes tighter as the speed increases or the distance decreases, hence we expect better temporal accuracy for fast movement or close interaction. In practice we calculate global minima by finding the point with minimum distance, and local minima by finding points nearer than its immediate neighbours.

We have investigated several ways of selecting minima. One method is to calculate the closest point on the trajectory to each object. A disadvantage of this is that some of the minima will correspond to objects which are far away and hence of little interest. These points will affect the later distributions so are best disregarded. We overcome this by calculating the distance to the *nearest* object at each point on the trajectory. A plot of the distance to the nearest vehicle for the trajectory in figure 1 is shown in figure **??**. The sharp points at the top of this plot correspond to when we are midway between two objects. The parabola lower down are the local minima which we use as our landmark points. We are likely to have several local minima for each object, and if the speed is low measurement noise may introducing extra minima. We can eliminate some of these minima by applying some smoothing to the curve, but this is still likely to leave more than one minima per object. This may be useful in capturing the length of time two objects are close or whether they have repeated interactions. Here we have opted to just choose at most one minima per object, actually the closest point.

3. Assessing the likelihood of a single interaction

We now have a sequence of landmark points on each trajectory. At each landmark point we can measure two quantities: speed and distance. We now build up a distribution of these measurements taken from all the landmarks on the paths in our sample set. This distribution is then used to assign a probability to each event. In other applications we could use other quantities such as curvature to further characterise the events.

Figure **??** shows the values of speed and distance for all the landmark points in our sample set of trajectories. From this diagram we build up a plot of the cumulative probability, i.e. we calculate the probability of a particular event with

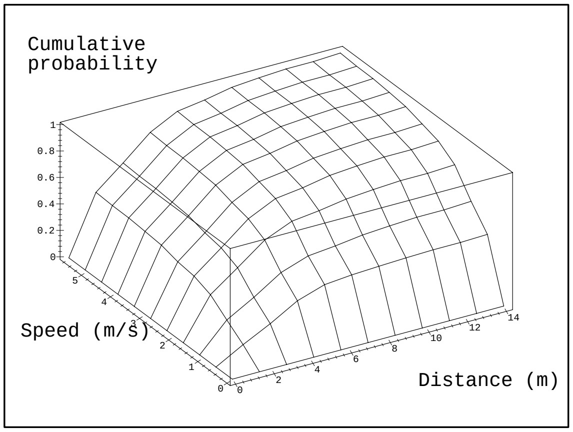

Figure 4. The cumulative probabilities for speed and distance

speed s and distance d *or worse* happening, i.e. an event with speed less than or equal to s and distance less than of equal to d. This is show in figure **??**. The cumulative probability can be calculated by simply counting the number of points with speed less than s and distance less than d. We then arrive at a probability by simply dividing by the number of points in the distribution. As there are a good number of minima, even from a small number of trajectories, directly counting the values will give us a good estimate of the probabilities and avoid the problems of trying to fit a function to the data.

For the car park scenario these probabilities fit with our intuition that we require both speed and distance to be low for a suspicious occurrence. Someone moving quickly along the edge of a line of cars will not be regarded as unusual. Likewise standing still a long way away from any cars is not considered unusual.

4. Un-ordered sequences

Assigning probabilities to the landmarks gives an ordered sequence of probabilities. There are many possible routes that an object can take through the scene. These depend on its final destination and also the position of the other objects. Therefore there are many possible sequences, and a particular sequence is unlikely to be repeated often, making it difficult to classify these sequences. Ideally we would like to obtain a representation of a trajectory which does not depend on the arrangement of objects. In particular we would like one that does not depend on the order in which the events occur.

Suppose we have an ordered sequence $(a_1, a_2, \ldots, a_n)$ of probabilities. We would like to make all possible permutations of this sequence equivalent, i.e.

$$\begin{aligned}(a_1, a_2, \ldots, a_n) &\sim (a_2, a_1, a_3, \ldots, a_n) \\ &\sim (a_n, a_{n-1}, \ldots, a_1).\end{aligned}$$

In mathematical terms we do this by considering the quotient space $\mathbf{R}^n/\Sigma_n$, where Σ_n is the group of all possible permutations on n elements. Unfortunately this is an abstract space which is difficult to represent on a computer. There are also problems with comparing elements in such a space.

Here we produce an unordered representation by simply sorting each sequence in terms of increasing probability. This gives us monotonically increasing sequences. We can compare two such sorted sequences in a term by term fashion. In effect this compares the two most interesting (lowest probability) items in each sequence, and then the second most interesting items and so on. This fits well with our application. The representation also allows us to compare sequences of different lengths.

5. Final Analysis

The final stage in the algorithm is to characterise monotonically increasing sequences of probabilities.

In the car-park scenario we expect people to get out of a car and walk towards an exit. This would give us a diagram similar to those in figure **??**. The first two probabilities are low corresponding to when the person gets out of the car and interacts with that car and the neighbouring vehicle. The probabilities then rapidly increase as all the other interactions happen at a reasonable speed and distance. We would expect a similar diagram if the person enters the scene and walks towards a particular car.

For atypical behaviour we expect a person to move slowly close to a number of cars. This will be indicated by a large number of events with low probability as shown in figure **??**. Note how this does not depend on the particular route taken through the scene.

There are a number of ways we can interpret this data, given that that the final results we want is a simple normal/atypical classification. We could, for example, choose the fifth, or tenth, item in the sequence and compare the probabilities, flagging low values. Alternatively we can choose a threshold for the probabilities, say 0.2, and count the number of events which lie below this value. It may be possible to use an un-supervised learning technique similar to [**?**] to model the distribution of these sequences.

Here we use a supervised learning technique to classify each sequence in our data set into two classes: *normal* and *atypical*. A weighted sum, $\sum a_i p_i$, of the first five probabilities, p_i, in each sequence is used. If the sum is less than

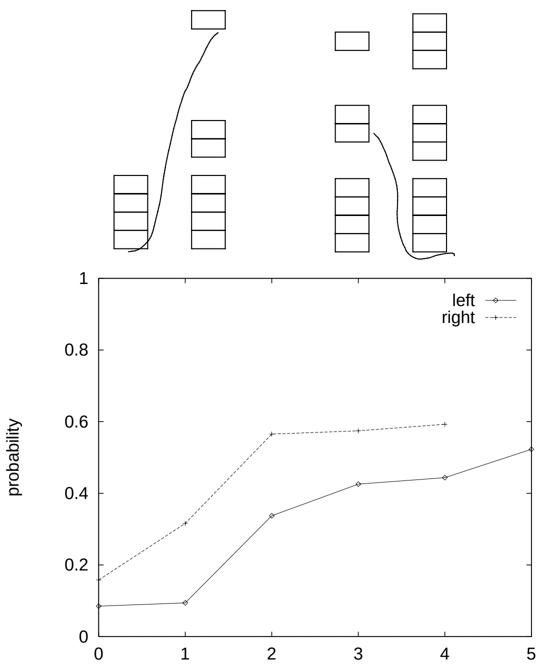

Figure 5. Two normal trajectories and their sequences of probabilities. The squares on the left-hand side represent the positions of vehicles.

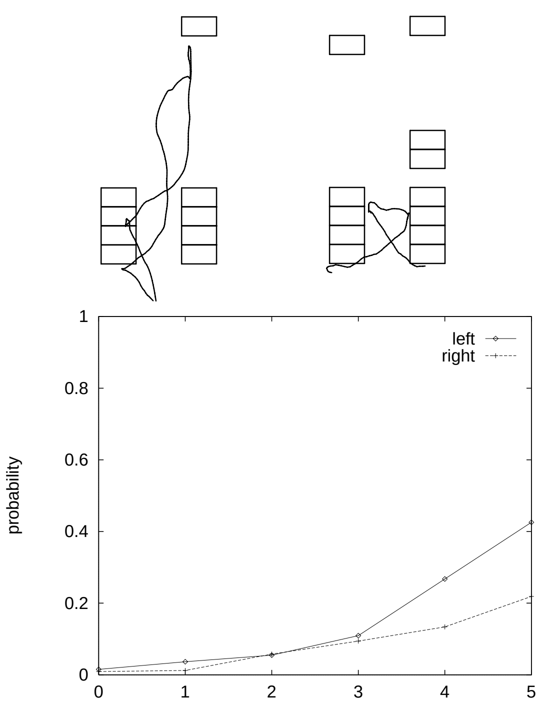

Figure 6. Two suspicious trajectories and their sequences of probabilities

0.5 then the trajectory is classified as being atypical. Thirty three of the trajectories including four atypical ones are used as our training set.

In the training phase we take an exhaustive search over all possible weights, each weight taking values between 0 and 1. We select those weights which correctly classify the most trajectories. We have found that a large set of different weights will all produce a good classification with only a few miss-classifications. This large set of possible weights was expected as visually there is a clear difference in the sequence plots between the two classes. For the testing stage we take all the sets of weights with three or fewer miss-classifications and calculate the median value for each weight. The remaining 26 trajectories were then classified using these weights. This correctly identified the one atypical trajectory and only yielded one false positive.

6. Conclusion

We have presented a framework for analysing sequence of interactions. For different scenarios it should be possible to alter the individual stages. For example by using a different set of landmarks or by using a different criteria for the likelihood of each landmark.

In further work we hope to extend this framework to capture such features as repeated interactions with the same car and also the time spent close to any particular vehicle.

References

[1] A. Baumberg and D. C. Hogg. An efficient method for countor tracking using active shape models. In *Proceedings of the IEEE Workshop on Motion on Non-rigid and Articulated Objects*, pages 194–199. IEEE Press, 1994.

[2] A. Baumberg and D. C. Hogg. Learning flexiable models from image sequences. In *Proceedings of the European Conference on Computer Vision*, volume 1, pages 299–308, 1994.

[3] A. Baumberg and D. C. Hogg. Learning spationtemporal models from examples. *Image and Vision Computing special issue BMVC9*, 1996.

[4] F. L. Bookstein. *Morphometric Tools for Landmark Data*. Cambridge University Press, 1991.

[5] F. L. Bookstein. Landmark methods for forms without landmarks: Localizing group differences in outline shape. *Medical Image Analysis 1*, pages 225–243, 1997.

[6] H. Buxton and S. Gong. Visual suveillance in a dynamic and uncertain world. *Artifical Intelligence*, 78:431–459, 1995.

[7] N. Johnson and D. C. Hogg. Learning the distribution of object trajectories for event recognition. *Image and Vision Computing*, 14:609–615, 1996.

[8] M. Monhaupt and B. Neumann. On the use of motion concepts for top-down control in trafic scenes. In *Proc. ECCV*, pages 598–600, 1990.

[9] H. H. Nagel. From image sequence towards conceptual descriptions. *Image and Vision Computing*, 6(2):59–74, 1988.

[10] P. Remagnino, A. Baumberg, T. Grove, D. Hogg, T. Tan, A. Worrall, and K. Baker. An intergrated traffic and pedestrian model-bassed vision system. In *Proc. BMVC*, pages 380–389, 1997.

[11] P. Remagnino, T. Tan, and K. Baker. Agent orientated annotation in model based visual surveillance. To be presented at ICCV '98.

[12] T. N. Tan, G. D. Sullivan, K, and D. Baker. Recognising objects on the ground-plane. *Image and Vision Computing*, 12:164–172, 1994.