International Journal of Computer Vision 37(2), 209–215, 2000

Statistical Models of Object Interaction

R.J. MORRIS
Department of Statistics, University of Leeds, LS2 9JT, England
rjm@amsta.leeds.ac.uk

D.C. HOGG
School of Computer Studies, University of Leeds, LS2 9JT, England
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 based on the speed and distance. Sequences of such probabilities are then sorted in increasing order. Finally a weighted sum of the first few elements in this weighted list is used to classify trajectories in a supervised learning framework.

Keywords: behaviour modelling, object interactions, landmark data

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 prerequisite for the work undertaken, two model-based systems are used to track the motion of people and vehicles. The first (Baumberg and Hogg, 1994a, 1994b, 1996) uses active shape models to track non-rigid objects, in our case people. The second (Tan et al., 1994) uses geometric 3D models to track rigid objects: cars. These systems have been integrated to handle mutual occlusion (Remagnino et al., 1997). Both systems provide coordinates on the ground plane of the objects being tracked. It is this information that we use as our input data. For this experiment we used the person tracking component but the positions of the cars were estimated by hand. For our purposes this is a reasonable approximation. In the final system we would use car positions automatically generated by the vehicle tracker.

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 (Monhaupt and Neumann, 1990; Nagel, 1988). Probabilistic or Bayesian networks have been used to help construct such descriptions (Buxton and Gong, 1995; Remagnino et al., 1997). Hidden Markov Models can be used to represent changes between states which can then be used to classify particular patterns of actions (Bobick and Wilson, 1995; Starner and Pentland, 1995; Bregler, 1997). The problem here is that the full network to represent all the possible paths would be quite large and without a large amount of training data, the estimation

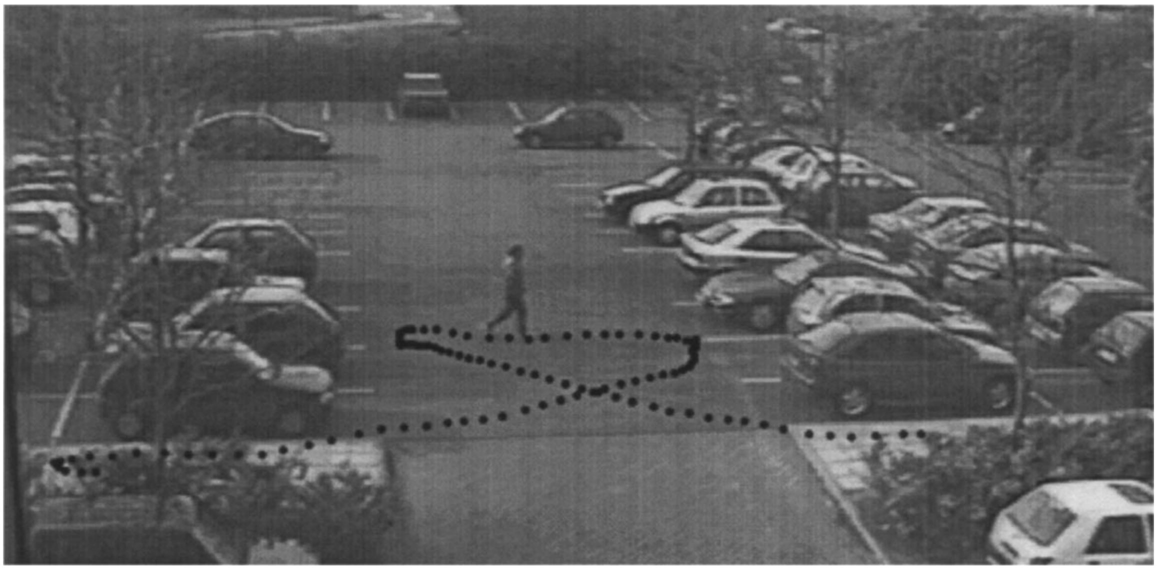

Figure 1. The car park scene and one of the trajectories through it.

of transition probabilities would be unreliable. There is also a question about how such a system would respond to a trajectory which had not been presented to it before, which is what we expect to happen in our situation.

An alternative approach (Johnson and Hogg, 1996) 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.

2. Defining the Problem

Figure 1 shows the car park scene used in our experiments and one of the trajectories through it. In all 129 trajectories were collected from about two hours of footage of the morning rush hour. This data includes 11 atypical trajectories which were provided by actors to test the system.

Figure 2 shows a selection of some of the paths which can occur. In the figure we have a top down view of the car park with cars shown as small boxes and the trajectories of people shown as curves. There are some patterns in the data, for instance the top row shows people walking along a path at the front of the car park. However if we look at the rest of the trajectories there are no such obvious patterns. In the second and third rows of the figure are trajectories of people exhibiting normal behaviour. They typically start near a particular vehicle and proceed in a reasonably straight path towards an exit. The routes they take can vary considerably, depending on where they leave and enter the scene, the cars they are walking away from and the arrangement of other cars in the scene. Along the way they may pass

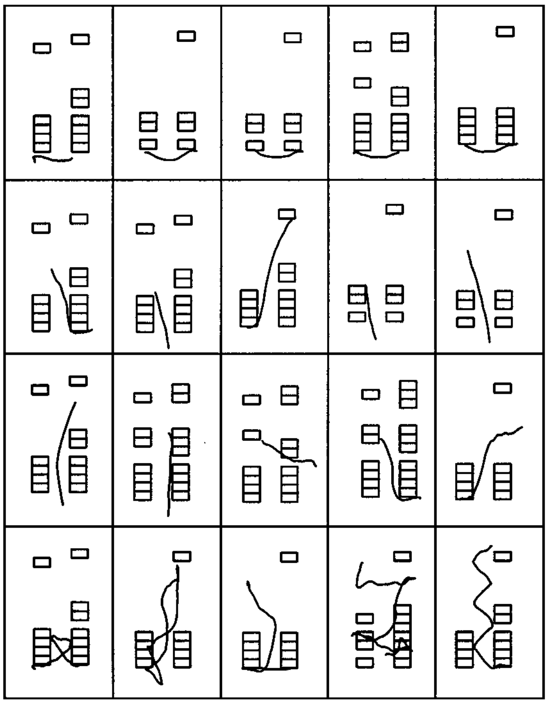

Figure 2. Some of the trajectories through the car park.

other cars, but normally at some speed, only coming to rest at their final destination. The bottom row shows some more atypical trajectories. In these the path is less direct, and the subject often slows down or stops near several vehicles.

The main task addressed in this paper is to produce a representation of the trajectories which encapsulates the relevant variability in the data allowing us to identify atypical behaviours. 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 eliminates most of the 'noise'. Once we have such a representation we can use standard statistical techniques to classify the trajectories. A method such as Johnson-Hogg's (Johnson and Hogg, 1996), where a distribution of the most likely trajectories is built up and then used to classify individual paths, will not work here as most trajectories will not occur more than once so almost all trajectories will be classed as atypical.

The approach we follow here is to identify individual events which capture the most interesting points on a trajectory, and examine the types and combinations of such events. Here we use the point of closest approach to each vehicle as discussed in Section 3. The events are described by considering the speed of the person and distance between the person and the vehicle. From such a description we can deduce how likely each event is; this is discussed in Section 4. This gives us an ordered sequence of probabilities. There is still a large variation in the sequences which can occur and these depend heavily on the arrangement of cars in the scene. We can factor out this information by sorting the sequence in terms of increasing probability (Section 5). The resulting sequences easily distinguish between typical and atypical interactions and the final section describes how we use a simple supervised learning technique to classify them.

The motivation behind the algorithm is worth some comment. Previous work examined the ratio of the length of the curve to the distance between its end points. This neatly summed up the statement "people tend to walk in straight lines". This suggested that the way to include the interactions with objects was to devise a simple but precise statement in English to describe types of behaviour we are trying to capture. Such a statement which captured the interactions with objects is that "moving slowly near to a number of cars is unusual". Once we had such a statement it was a relatively easy matter to devise an algorithm to find such events.

3. 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 (Bookstein, 1991) particularly in the medical field. In other applications various geometric features of a curve such as the vertices or inflections can be used. For example Bookstein (Bookstein, 1997) has recently developed a novel statistical approach which uses bending energy of interpolating splines to specify landmarks. 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.

If we consider the trajectory shown in Fig. 1 and calculate the minimum distance from each point on the trajectory to any of the cars we get the plot shown in Fig. 3. The sharp points at the top of this plot correspond to when we are midway between two objects. The most striking feature about this graph are the minima, which suggests that they are good candidates for our landmark points.

There are several ways in which we can select minima:

1. For each object find the closest point on that trajectory to that object. This gives us one landmark for each object.
2. Consider the function

$$f(t) = \min_i(\text{distance to object } i \text{ at time } t),$$

and find all local minima of this function.

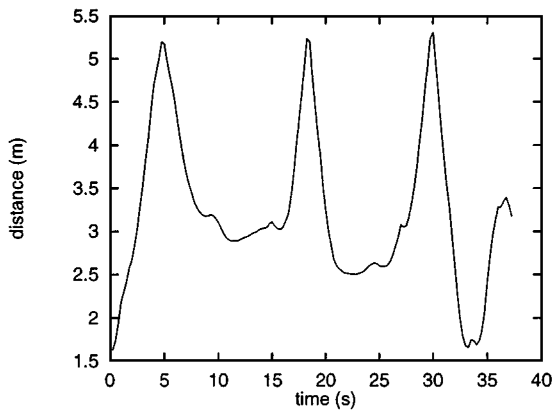

Figure 3. The distance to the currently nearest car.

3. Take the intersections of 1. and 2. i.e. find the global minimum of distance to each object, but reject those where some other object is closer.

In the first method some of the minima will correspond to objects which are far away and of little interest, but they will still affect the characteristics of later distributions, in particular making it sensitive to the number of object in the scene. In the second method measurement noise is a major problems. If the speed is low there may be several local minima for each object. We can eliminate some of these minima by smoothing 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 for the third method, which combines the good features of the first two.

Some pre-processing is carried out prior to calculating minima. First very short trajectories are removed. These are occasionally generated by tracker error. Then the trajectories are smoothed by averaging over a 1 second interval.

4. Assessing the Likelihood of a Single Interaction

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

Figure 4 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 histogram of the cumulative probabilities. For an event with speed s and distance d we calculate the cumulative probability of an events occurring with speed $\leq s$ and distance $\leq d$. This is show in Fig. 5. 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, from even a small number of trajectories, directly counting the values will give us a

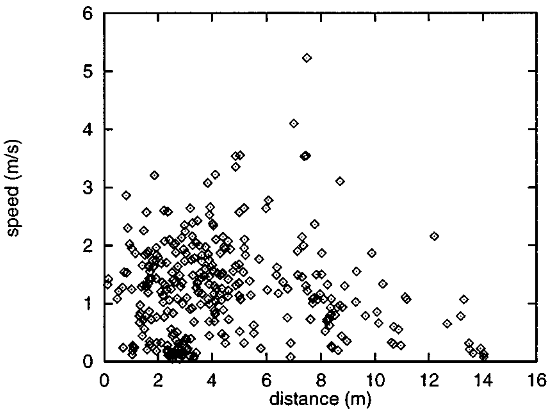

Figure 4. The distribution of speed and distance.

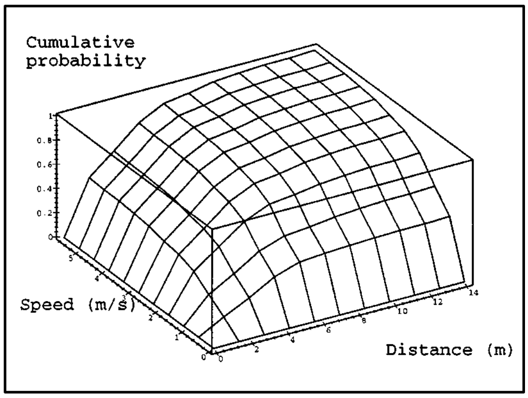

Figure 5. The cumulative probabilities for speed and distance.

good estimate of the probabilities. Fitting a Gaussian mixtures and then calculating cumulative probabilities would give much the same results as our method, and may prove to be a better technique in the long run. One complication with such a method is that the distribution is constrained to lie in $s \geq 0$, $t \geq 0$. Fitting Gaussians would give non zero probabilities for negative value of s or t.

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 from any cars is not considered unusual.

Using probabilities in this manner maintains a continuous representation of the data, contrasting with discrete approaches which introduce arbitrary thresholds to classify event into particular states. The method also eliminates the need for any parameters to be set and is self-calibrating. For a given scene, the algorithm will learn what counts as a slow movement by comparing it with the sample.

Rather than using cumulative probabilities we could approximate the distribution by a Gaussian mixture and then calculate the probability density at each landmark. This gives a very different representation of the data. Two landmarks one with low speed and one with high speed will have very different cumulative probabilities but may have similar probability densities. Ideally we would like to distinguish between such landmarks. Furthermore if we examine Fig. 4 we see that there is a cluster of landmarks with speed ≤ 0.5 m/s and distances between 2 and 4 meters. Points in this cluster would have high probability densities but low cumulative probabilities. For our purposes we would like to classify such points as unusual and this motivates our choice of representation.

5. 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.

A natural extension of this approach is to use *Order Statistics* and examine the distributions of the i-th element in the sorted list. See (Gumbel, 1958) for a detailed analysis of this technique.

We can think of this step as representing the grammar of the system. We have adopted a very loose grammar, where many different sequences are though of as being equivalent. This contrasts with the strict grammar of the Johnson-Hogg system, and the more involved grammars in the Bayesian network type approaches. Our approach illustrates how little information is needed to get meaningful results. It should also be regarded as just a first approximation. There are several types of events which it does not capture: for instance repeated approaches to the same vehicle. However such events are rare and for 90 percent of trajectories, our model is good enough. We are looking at ways to expand our approach to include a richer grammar.

6. 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 Fig. 6. 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 Fig. 7. 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 result 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.

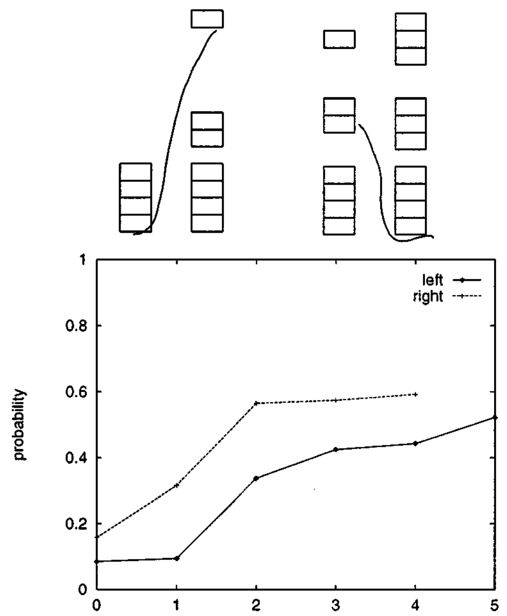

Figure 6. Two normal trajectories and their sequences of probabilities.

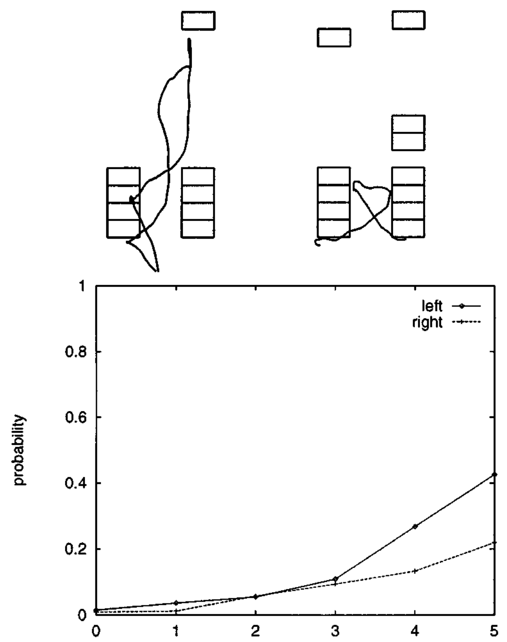

Figure 7. Two suspicious trajectories and their sequences of probabilities.

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 unsupervised learning technique similar to (Johnson and Hogg, 1996) or (Bulpitt, 1995) to model the distribution of these sequences.

Here we use a planer decision surface to distinguish between normal and atypical behaviour. Let p_i be the first five probabilities in a sequence and let a_i be a set of weights. If there are fewer than five probabilities in the sequence then the remaining probabilities are set to 1. We use the classification rule

$$\text{IF} \sum a_i \, p_i > 0.5 \text{ THEN atypical ELSE normal}$$

to distinguish trajectories. In the training phase we find the set of weights which best classify the trajectories, and we then use these weights in the testing phase. In practice the weights were found by taking an exhaustive search over all possible weights. Other techniques like discriminant analysis could also be used here but as the problem was fairly simple we felt it did not justify a more advanced technique.

The training data consisted of 59 trajectories, which included five atypical trajectories. We found that a large set of different weights all produced a good classification of the data with only a few miss-classifications. We selected those sets of weights which gave four or fewer mis-classifications and then calculated the median value for each weights.

The test set consisted of 70 trajectories including 6 atypical ones. Using the mean weights 10 trajectories were classified as atypical including all 6 atypical ones. This represents only 4 false classifications in 45 minutes of footage.

7. Conclusion

We have presented a framework for analysing sequence of interactions, which has shown good results at classifying results in a car park application. 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.

The project as a whole represents a full image understanding solution, from identifying objects in the scene to interpreting behaviour and answering high level questions about what is happening. It would probably work best in conjunction with a human operator. For instance sounding an alarm to draw the attention of an operator, when a possibly suspicious event is detected.

Currently the algorithm incorporates a lot of prior knowledge about the types of behaviour we are trying to detect. For example the assumption that low speed and low distance are suspicious. We are currently exploring ways in which we can remove some of these assumptions. In further work we also hope to extend the framework to capture repeated interactions with the same car and also the time spent close to any particular vehicle.

More work is also needed in obtaining a better data set. Problems with the tracker mean its difficult to track people weaving in and out of the vehicles. In collecting the data we tried to avoid such situations. Hence the atypical trajectories do not really represent the possible trajectories of car thieves.

Acknowledgment

This work was supported as part of the EPSRC IMV project, Grant No. GR/K46620, while the first author was at the School of Computer Science, University of Leeds.

References

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

Baumberg, A. and Hogg, D.C. 1994b. Learning flexiable models from image sequences. In *Proceedings of the European Conference on Computer Vision*, Vol. 1, pp. 299–308.

Baumberg, A. and Hogg, D.C. 1996. Learning spationtemporal models from examples. *Image and Vision Computing*, 9(special issue BMVC).

Bobick, A.F. and Wilson, A.D. 1995. A state-based technique for the summarisation and recognition of gesture. In *Proc. 5th Int. Conf. on Computer Vision*, pp. 382–388.

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

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

Bregler, C. 1997. Learning and recognizing human dynamics in video sequences. In *Proc. Conf. Computer Vision and Pattern Recognition*, pp. 568–574.

Bulpitt, A.J. and Allinson, N.M. 1993. Human Motion Recognition Using Co-operative ART Networks. In *Proc. World Congress on Neural Networks*, Vol. 3, pp. 708–711.

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

Gumbel, E.J. 1958. *Statistics of Extremes*. Columbia University Press.

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

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

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

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

Remagnino, P., Tan, T., and Baker, K. 1998. Agent orientated annotation in model based visual surveillance. In *Proc. ICCV*, pp. 857–862.

Starner, T. and Pentland, A. 1995. Visual recognition of american sign language using hidden markov models. In *Proc. Int. Workshop on Automatic Face and Gesture Recognition*.

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