Active Shape Model Search using Pairwise Geometric Histograms

E.C. Di Mauro, T.F. Cootes, C.J. Taylor and A. Lanitis,

Department of Medical Biophysics,
University of Manchester, M13 9PT

email: ecd@sv1.smb.man.ac.uk

Abstract

We have investigated a model based approach to interpreting 3D outdoor scenes containing motor cars. We use a 2D statistical shape model which can deal with both intrinsic variability arising from different designs of car and extrinsic variability arising from changes (over a limited range) in 3D viewpoint. The aim was to use Active Shape Models (ASM) search to locate cars in images. However, because of the changing environment the existing methods of using image evidence were inappropriate and ASM search was unreliable. We describe a method of labelling each edge segment in the scene with the likelihood that it corresponds to each model edge segment, using local geometrical context encoded in a set of Pairwise Geometric Histograms (PGHs). The labels are used in a modified ASM search algorithm which we show, using both synthetic and real data, to be much more robust than either simple edge based or grey−level model based ASM search.

1 Introduction

We are interested in finding motor cars in cluttered outdoor scenes. Because different models of car have different shapes but there is a common class structure, we have investigated the use of statistical shape models [1] to create a generic car model. Unlike Worral *et al.* [2] who use 3D statistical models, we have incorporated the apparent change in shape, resulting from change in viewpoint over a limited range, into a 2D model. Our intention is to use a set of 2D models each dealing with a limited portion of the viewsphere. The ASM framework is necessary in this domain, because it provides the statistical boundary for the allowed variability of the model. When we began to implement this scheme, we found that the shape modelling strategy was straightforward (and not reported in detail here) but that Active Shape Model (ASM) search [1], which has proved very robust in other applications, was unreliable with these images. If the search was conducted using simple edge evidence to refine the pose and shape of the model, it would often converge to a plausible but incorrect solution. If local grey−level models were used to drive the refinement process [3] – an approach which has proved effective in other applications − there was no improvement because the constantly changing environment rendered the models ineffective. To deal with these problems we have developed a new approach. The basic idea is to use an edge based ASM framework, but to make the search more directed by labelling line segments in both the model and the image with their local geometric

BMVC 1996 doi:10.5244/C.10.13

context. This ensures that model features are preferentially attracted, during model refinement, to scene features which they are likely to match. The geometrical context is encoded in a set of Pairwise Geometric Histograms (PGH) [4]. We show, using both synthetic and real images, that ASM search using the new method is much more robust than either simple edge based or grey–level model based search. The edges that belong to the car do not always appear in the images, because of uneven lighting or shape smoothness. By having a constraint on the allowed general variability, location uncertainty due to lack of data is reduced.

2 Background

2.1 Active Shape Models

Active Shape Models (ASMs) [1,3] provide a useful means for locating objects in images. An ASM relies on having a statistical model of the expected shape of an object, representing the object by a set of model points. Associated with each model point is a model of the image evidence expected at that point. This can either be an edge model or a more general statistical model of the grey–level appearance of the region around the point. Both shape models and (where necessary) grey–level models are generated from sets of training images. To search for an object using an Active Shape Model an initial estimate of the solution is iteratively refined. At each step the image model associated with each point is run along a normal through its current position, and the best match is found. The model parameters are then updated to move the model points toward these suggested positions. The constraints imposed by the shape model help to ensure rapid and robust convergence.

2.2 Limitations of ASMs

In situations where the grey–level appearance of an object is reasonably predictable (such as in medical images or constrained industrial inspection problems) ASMs using grey–level models at each point have proved extremely useful [1, 3, 5]. The advantage of using a grey–level statistical model for the region around each point is the ability to provide a reliable candidate position for the point along the search profile. However, in some domains, such as outdoor scenes, the grey–level appearance can change significantly and unpredictably. In such cases the grey–level models fail. The most reliable features in such scenes are edges. However, when we search along a profile for edges, it is common to find several candidates and it is not clear which is the most suitable to aim for. Thus in cluttered scenes (or when the model itself has several nearby edges) ASM search does not give satisfactory results, unless initialised very close to the true position (see Fig. 1). If each edge found in an image could be labelled with its local context, and each model point was tuned to locate edges with a particular context, this would help to overcome the ambiguities. Pairwise Geometric Histograms provide a mechanism for contextually labelling edge segments and might form a basis for robust edge based ASM search.

2.3 Pairwise Geometric Histograms

Pairwise Geometric Histograms were first introduced by Evans, Thacker and Mayhew [4]. They sought to represent shape by recording the distribution of pairwise geometrical relationships between local shape features. The problem they were addressing was the recognition and location of multiple rigid objects within an image.

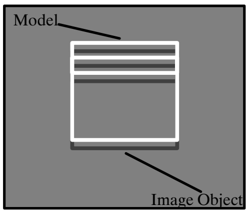

Fig. 1 − When searching for edges an ASM can get stuck in a local minima when the wrong edges are located. Similar problems occur if there is much clutter in the image.

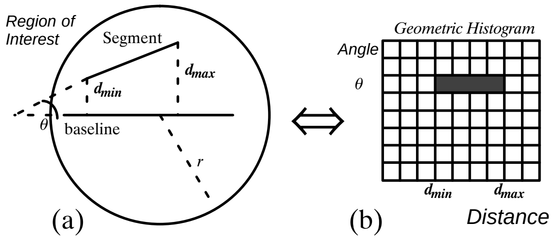

Fig. 2 − (a) Angle between baseline and segment and distances between baseline and segment. (b)Corresponding entry in the angle−distance Geometric Histogram.

The method had to cope with changes in lighting and partial occlusions of the object. The proposed solution had two parts: a representational scheme to model the objects and an algorithm to match corresponding object and model features. Each shape was represented by a set of straight lines approximating the boundary of the object. For each segment in the boundary, a histogram was built in the following way. Each segment was in turn called a “baseline”. Around each baseline a circle was drawn of radius r (Fig. 2). All the other segments occurring within this circle of interest were considered to form a set of local shape features of the object. The geometric relationship between the baseline and each local segment was represented by three parameters: the angle θ between them and the range of perpendicular distances from the end−points of the segment to the baseline. The representational scheme was then based on recording the distribution of values of these geometrical features in an angle−distance Geometric Histogram (Fig. 2(b)). Each entry in the histogram was given a value equal to the product of the length of the baseline times the length of the relevant segment. Uncertainty in position and orientation was modelled by blurring the entries in the histogram along each axis. This process was applied to all the n segments in the object's boundary, creating a set of n model histograms. The same process was then applied to the test image, producing a larger set of m histograms. In general, $m>n$, because other objects and noise are present in the test image. The histogram for each model line was compared with each one of the m histograms from the test image − a total of $m \times n$ comparisons. If each histogram is considered to be a vector then the degree of match between the histogram M_j of the jth model segment and the histogram I of an image segment is given by the dot product correlation D_j:

$$D_j = \sum_i \sqrt{I_i \, M_{ij}} \quad (1)$$

where the summation is over the elements of the histogram. This is related to the maximum likelihood similarity metric and can therefore be considered a general result for comparing two probability distributions. The model feature with the highest correlation value was finally taken as a match for the particular image feature. The pose of the object could then be determined.

3 Active Shape Models using Pairwise Geometric Histograms

In order to use PGHs in the ASM search, we break down the model into a number of line segments. These are simply the straight lines between model points. For a given model line segment we generate a PGH from each training image, then merge the

PGHs together to give a mean PGH for that segment. Given a new image we pre–process it to first extract all line segments, then we calculate the PGH for each segment. When searching with an ASM we associate each model point with the model line segment on which it lies. To search for a suggested new position for a point we look for intersections of a normal through the point, with image segment lines, and select the intersection point which is associated with the image line whose PGH best matches the PGH of the model segment.

3.1 Building the edge segment images

The original raw image undergoes a set of pre–processing stages as shown in Fig. 3. The edges are first extracted using a Canny filter [6]. The value of σ (in pixels) for the Canny filter is chosen to be reasonably large (from 6 to 10) in order to reduce the number of edges recovered (thus limiting the effect of noise and providing a schematic of the object shape). The image is binarised, to reject the effect of uneven lighting of the object features (it is the edge presence that is important not the edge strength). The edges are thinned to make sure that multiple edge responses are treated as one combined edge. The edges consisting of a single pixel are removed, simplifying the image further. Multiply connected edge segments are decomposed into separate sub–edges. Finally, a polygonal approximation algorithm is used to approximate the edges with straight line segments [7]. Although this procedure is relatively robust, changes in the lighting and pose of a moving 3D object can cause considerable variability in the number and quality of the edges in the pre–processed image. This affects search, because line segments can appear and disappear from consecutive images of a sequence and continuous edges can become fragmented. Evans *et al.* [4] considered such effects and showed that the PGH approach is robust in their presence.

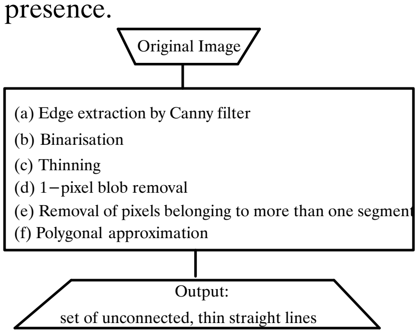

Fig. 3 – Pre–processing of the raw image to produce an image made up of a set of unconnected, thin straight lines.

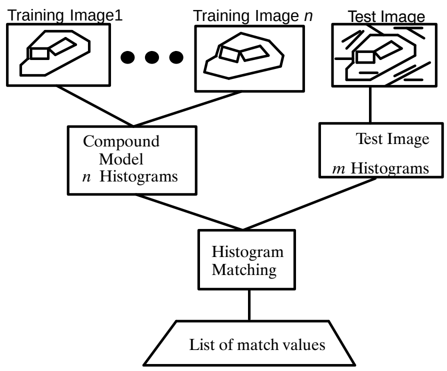

Fig. 4 – Labelling and matching algorithm.

3.2 Labelling the images

Once the scene has been approximated by a set of straight line segments, these are 'soft–labelled'. For each scene segment, the likelihood that it corresponds to each of the model segments is calculated. This approach is similar to that outlined by Evans *et al.* [4]. A set of histograms is built for the model and a set of histograms is built for the test image (Fig. 4) If the model contains n lines connecting the landmark points, then n histograms will be created, one for each baseline. The test images contain a variable number m of segments, where m is generally much larger than n. We have adapted Evans *et al.*'s method to deal more satisfactorily with variable objects − they were principally interested in rigid objects. We are interested in objects whose appearance changes because of intrinsic variability (e.g. a different car model) or extrinsic variability (e.g. different pose, lighting or perspective). Evans *et al.* blur each entry in the model histograms in order to cope with inaccuracies in angle and distance measurements. We include variability explicitly in the model histograms by summing the model histograms for the same baseline for all the images in the training set. This process produces models which are more specific and more statistically sound than arbitrary blurring each histogram model entry. The method of angle and distance measurement suggested by Evans *et al.* [4] gives rise to ambiguities because lines at either side of the baseline can be indistinguishable. If we assume that the baseline has a direction, we can project the end−points of a target line onto the normal to the baseline to give a signed distance from the line (Fig. 5b). Alternatively (Fig. 5c), we can measure distances to the end−points radially from the centre of the baseline (of course, this latter introduces different ambiguities and loses some of the desired properties of the representation but still gives better performance in our application). In conclusion, Table 1 lists the main differences between the standard PGH algorithm proposed by Evans *et al.*[4] and our modified version.

Standard PGH	***Modified PGH***
Rigid bodies	Variable shape bodies
2D objects	3D objects
Objects lie in a plane parallel to the image plane	Objects lie in any plane
Model histogram entries are blurred	Model histograms are created by summing histograms from a training data set
Normal distances only	Directed normal or radial distances

Table 1.− Differences between the Standard and the Modified PGH algorithms.

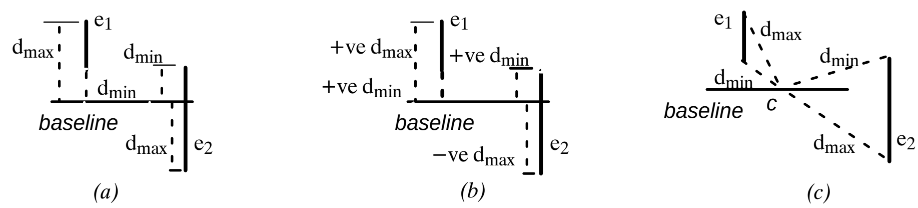

Fig. 5 − (a) shows two different edges being defined by identical values of normal−distance. (b) shows the same edges being discriminated by assigning a sign to the distances using the baseline as reference. (c) shows the same edges being discriminated by different radial−distances.

The choice of size of the Region Of Interest (ROI) used for constructing PGHs also affects the quality of matches. Histograms produced using evidence gathered over a large ROI are more specific because they are less likely to get a good match to a different pattern. However they also tend to include more noise and take longer to build, because they include data from more edges. The results in Table 1 show that the best choice for the ROI radius is to make it equal to the average length of the model's edges. In this way the scale relationship between edges is preserved and the whole baseline is always included within the ROI. Experiments with a variable ROI radius, selected to match the current baseline, produced low quality matches. Another problem which is associated with the choice of ROI is the choice of resolution of the histograms, because large ROI's require for higher number of bins in the histogram. Histogram binning is a major factor in the quality of match: high resolution in θ and d leads to more specific models, but also increases the computational complexity. Tests indicate that θ is more important than d. In fact the geometrical relationship between edges in an image is better captured by their angular relationship than by their distances. Distances are affected by scaling and zooming, whereas angles are invariant to both. Higher histogram specificity is thus achieved by increasing the resolution in θ. The results of systematic experiments are shown in Table 2. One hundred wire frame models of a turning car were analysed (with no noise present). The radius of the Region Of Interest, the resolution in angle θ and the resolution in distance of the histograms were given a range of values. The match values of the matches of the compound model edges with the edges of each example in the training set were recorded. The mean of all the fitness values across the training set and for all edges was then calculated. This provided the total rank values of Table 2. With reference to Table 2, the radius of the ROI, the resolution in angle θ and the resolution in distance of the histograms were given a range of values. The effect of these variations was measured as the rank of the pairwise matches for all the edges of the model in 100 different poses. A rank of 0 corresponds to a perfect match. From the table is evident that the best rank (0.47) is achieved when the histograms are given the highest resolution in θ (200), The lowest rank (15.8) is instead achieved when the resolution in θ is low. Variations in distance resolutions, on the other hand, do not produce dramatic changes in rank (compare col.s 6 & 7 and col.s 4 & 8). Radius has a limited effect on rank, however the highest rank is achieved when the radius is made equal to the average length in pixels of the edges of the car image, i.e. 100 (compare col.s 1, 2 & 3).

Col.no.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Rank	*6*	*4.7*	*8.3*	*15.8*	*13.4*	*2.3*	*2.04*	*13*	*0.47*
Radius	*200*	*100*	*50*	*50*	*50*	*100*	*100*	*100*	*100*
Angle	*50*	*50*	*50*	*10*	*20*	*100*	*100*	*10*	*200*
Distance	*50*	*50*	*50*	*10*	*20*	*100*	*10*	*100*	*10*

Table 2 – Tests for a synthetic model of a car.

3.3 Running the PGH driven Active Shape Models

Once each scene segment has been labelled with the likelihood that it corresponds to each of the model segments, a modified edge based ASM search can be performed. As in normal ASM search [1], an initial approximation is projected into the image and search lines are created at each model point, normal to the model boundary. Instead of attempting to move each model point towards the nearest edge, as would happen in simple edge based search, the likelihood values are inspected for each scene segment which intersects a search line. For each search line the intersection with the scene segment most likely to be the corresponding model segment is selected as the target position. Thus a model part is able to ignore nearby edge segments in favour of a more distant segment which is more likely to be the target segment (as evidenced by its local geometrical context). Once a set of target positions have been identified for the model points, the model parameters are updated in the normal way. This process is iterated until a stable solution is found. An efficient implementation is possible using a list of pre–sorted scene segments and calculating search–line/scene segment intersections analytically, rather than searching for intersections directly in the image.

4 Experiments

4.1 Experiments to determine the quality of PGH matching with synthetic images and models.

As a synthetic example we generated sets of 3D points representing the vertices of a hatchback car. Fig. 6 shows a typical set. When projected into an image (using a pinhole camera model) we arranged for its projection to be 200 pixels long, 100 pixels wide and 150 pixels high. The occluded points are hidden to the observer. The experiment we performed consisted of searching for the car amongst clutter segments with lengths comparable to the car model edges.

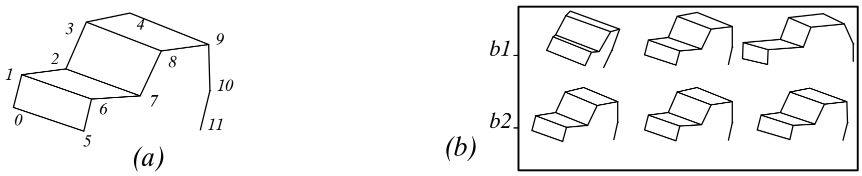

Fig. 6 – (a) shows the synthetic model of the hatchback car. Twelve landmark points are used. (b) shows the two PDM modes of variation of the turning car. The standard deviation is 3.

Two hundred edges of random length (from 5 pixels to 200 pixels), orientation and location were superimposed on the image of the model creating scene clutter (Fig. 7). Results were obtained comparing the performance of two search methods: the first uses the "nearest edge" method, the second uses contextual labelling of the edges. Tests performed on uncluttered images showed that both methods performed equally well. However when clutter was added, the "labelled edge" method out–performed the "nearest edge" method. Fig. 7 shows a set of results for different initial positions of the model. On the left Fig. 7(a) shows the models in their initial pose. In the middle (Fig. 7(b)) are the final results for the "nearest edge" search. On the right, Fig. 7(c) show the results for the "labelled edge" search. It is quite difficult for the reader to spot the target car model (in the centre of each image) amidst the clutter. The results were achieved by running 30 iterations of ASM, with search profiles 300 pixels long (approximately the length of the car). Fig. 7(i) shows results for an initial model rotated by 30 degrees and displaced of 100 pixels in x and y. The "nearest edge" result shows a shrunken car. The "labelled edge" result is correct. Fig. 7(ii) shows results for a slightly displaced initial position. The "nearest edge" result is good (except for the front part of the car), showing that the overall shape constraint of the ASM is effective. The "labelled edge" result is also correct. Fig. 7(iii) shows results for an initial model which is twice the size of the target. The "nearest edge" method fails dramatically, whereas the "labelled edge" method "shrinks" to fit the target. Fig. 7(iv) shows results

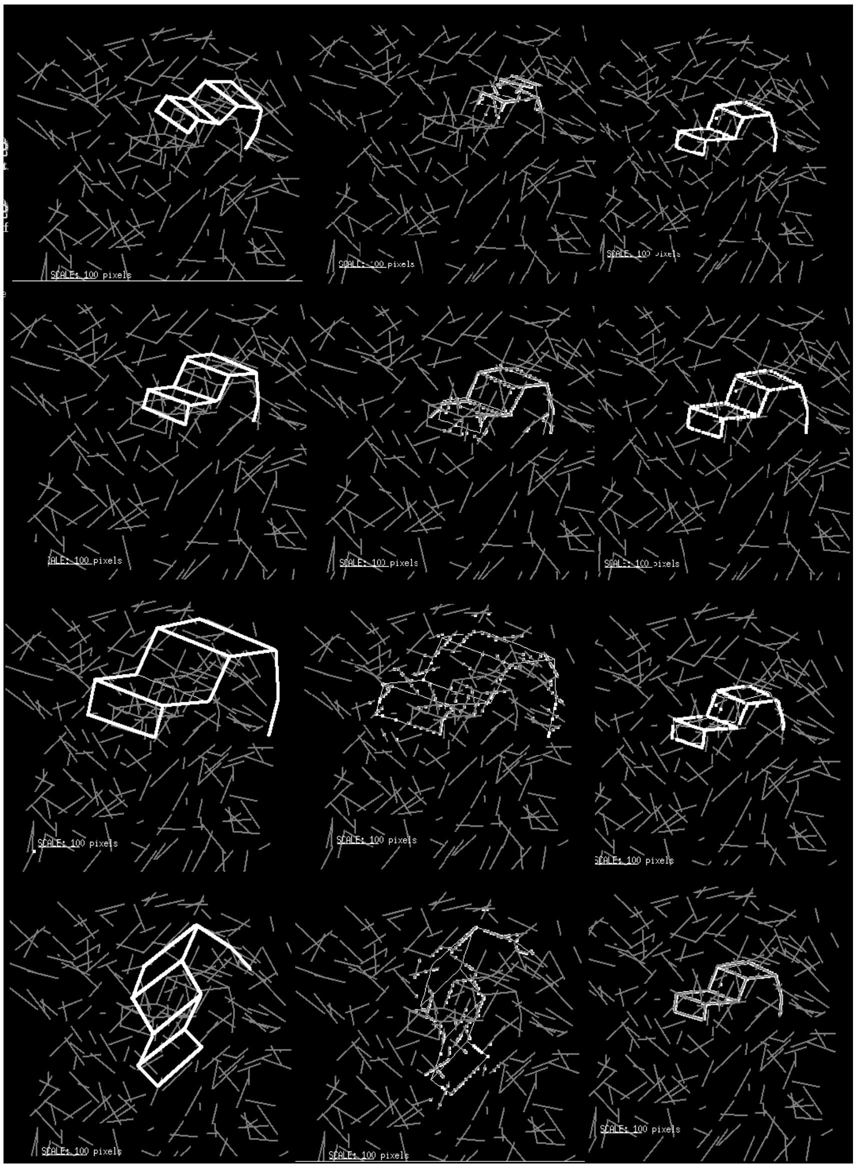

Fig. 7 – Comparison between "nearest edge" and "labelled edge" methods for tests with synthetic data. Column (a) shows initial model position (highlighted). Column (b) shows the final result for "nearest edge" method. Column (c) shows the final "labelled edge" result. For (b) and (c) bright blobs indicate point matches found.

for an initial model scaled by 150% and rotated of by 60 degrees. Only the "labelled edge" method succeeds in shrinking and rotating the model to fit the evidence.

4.2 Experiments Using Real Data

We have also performed experiments with real images. A hatchback car model was trained with images of car in different poses all taken from a similar viewpoint – see

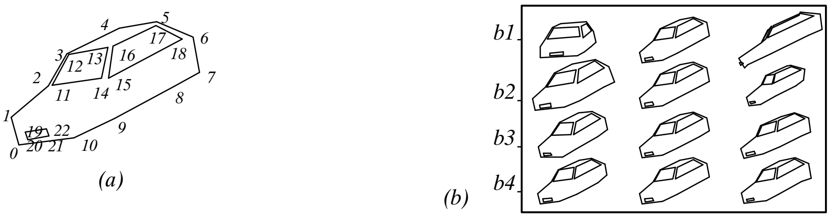

Fig. 8 – (a) shows a model of a real hatchback car. Twenty–three landmark points are used. (b) shows the first four PDM modes of variation of the turning car for a variation of ±3 standard deviations.

Fig. 8. The model was applied to a video sequence of a hatchback car turning a corner. The performance of the "labelled edge"method was compared with the two ASM methods currently available: the "strongest edge" method and the "grey–level matching" method, where the whole grey–level profile is matched. Fig. 9 shows the superior performance of the "labelled edge" method compared to the other two for the same car in a 3D scene.

(a) (c) (e)

(b) (d) (f)

Fig. 9 – Results for the search in a real 3D scene. (a) shows the edge detected image of the car. (b) shows a labelled edge image of the same car for a particular target edge (intensity indicates quality of match). (c) shows the starting model position. (d) shows the result for the "grey–level match" method. (e) shows the search result using the "labelled edge" method. (f) shows the result for the "strongest edge" method.

5 Discussion and conclusions

In this paper we have proposed a new method of searching for 3D objects in real outdoor scenes. Its main feature is that the search process is performed not on the original image, but on an image where all the significant edges have been extracted and labelled. The value of each label expresses the degree of confidence that the edge is a specific part of the target object. The ASM paradigm ensures that at each search iteration the model assumes only legal shapes. The labels are obtained by a matching mechanism based upon the Pairwise Geometric Histograms of Evans *et al.* [4]. We have, however, adapted the mechanism and extended it to variable objects in a 3D scene. The results in section 4 are encouraging, because the synthetic data tests suggest that the method is robust to severe noise effects. The tests with real images have shown that the the method provides a definite improvement in the search compared to either the grey–level based ASM or the strongest edge based ASM. The major problem, at present, lies in the correct choice of which edges of the object are specific enough to characterise the object but are still visible under variable lighting conditions and under variable perspective projection. Only consistent and repeatable labelling and image filtering can guarantee successful searches. At present we are carrying out extensive experiments to identify which parts of a car best lend themselves to this kind of directed search. Tracking is an easier problem once the car has been found in the initial frame of the sequence. We foresee that in a real–time tracking application image filtering and PGH histogram building would be performed only locally in the ROI, saving time and memory space.

Acknowledgements

The first and the last authors are funded by an EPSRC Project Grant. The second author is funded by an EPSRC Postdoctoral Fellowship.

References

[1] T.F.Cootes, C.J.Taylor, D.H.Cooper and J.Graham, Active Shape Models – Their Training and Application, *Computer Vision and Image Understanding* Vol. 61, No. 1, 1995. pp.38–59.

[2] Worral, A.D., Ferryman, J.M., Sullivan, G.D and Baker, K.D., Pose and Structure Recovery using Active Models, *Procs of the British Machine Vision Conference*,1995. pp.137–146.

[3] T.F. Cootes, C.J. Taylor and A. Lanitis. 'Active Shape Models: Evaluation of a Multi–Resolution Method For Improving Image Search'. *Procs. of the 5th British Machine Vision Conference 1994*, vol 1, 1994, pp. 327–336.

[4] Evans, A.C., Thacker, N.A. and Mayhew, J.E.W., The Use of Geometric Histograms for Model–Based Object Recognition. *Procs of the British Machine Vision Conference*,Vol. 2, 1993, pp. 429–438.

[5] S. Solloway, C.J. Taylor, C.E. Hutchinson and J.C. Waterton, Quantification of Articular Cartilage from MR Images Using Active Shape Models, *Procs of the European Conference on Computer Vision*, Springer, 1996, pp. 400–411.

[6] Canny, J, A Computational Approach to Edge Detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, Vol. PAMI–8, No. 6, Nov 1986, pp. 679–697.

[7] Wall, K. and Danielsson, P.E., A fast sequential method for polygonal approximation of digitized curve, Computer Vision, Graphics and Image Processing, Vol. 28, 1984, pp. 220–227.