Alvey MMI-007 Vehicle Exemplar :

Evaluation and Verification of Model Instances

Kay Brisdon

Deptartment of Computer Science
University of Reading, RG6 2AX

Introduction

It has proved very difficult to recognise three-dimensional objects in natural scenes, based only on two-dimensional features extracted from an image using bottom up methods. One of the major problems is that of relating the two-dimensional information available in the image to three-dimensional entities. This problem can be tackled by attempting to reconstruct knowledge about depth, and hence gain 2½-D information to work from, this method however is only applicable when fairly precise data, i.e. low-noise images, or multiple image views are available. Alternatively, viewpoint independent features in the image can be identified and used for matching to models. This is simpler but again restricted. However in the approach described here, using a model-based hypothesise-and-test strategy, hypothesised two-dimensional instances of three-dimensional models can be used to match against directly observable, view-specific image features. This obviates the necessity of obtaining three-dimensional data from the image, and as Lowe points out, viewpoint dependent matching is very much more powerful than any viewpoint invariant method [Lowe 1987].

This paper deals with the "test" part of the strategy. It assumes the image has already been sufficiently processed to produce an hypothesised instance of a known object [See Godden 1987, Morton 1987, Hutber 1987]. What is then required is a method for quantifying the acceptability of this hypothesis.

The Verification Process

Verification is a top-down process, which uses a geometric description of the object of interest to predict its two-dimensional appearance, and thereby compare it with the image data. This allows very precise tests to be carried out, as exact feature relationships are known. A novelty of this system is that the hypothesis is tested directly on the image data, rather than working in a symbolic domain. Constructing smoothed, segmented image curves from the output of a data-driven operator although an acknowledged weak spot in present vision systems, is the representation relied upon by many in object verification [Lowe 1987, Brooks 1984].

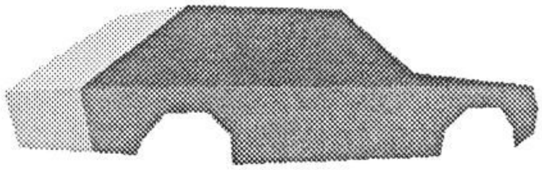
Shirai provided an example of the idea of returning to the original image and of using previously gained knowledge to guide the recognition process [Shirai 1978]. He detected and grouped clusters of similar intensity changes in an image, initially with a high threshold criterion, resulting in a series of conspicuous edge fragments. These he then classified as straight lines or ellipses, and matched to the program's object models. Partial matches resulting from this process were used to predict "missing" features in the image. These could be verified by then searching for specific edge evidence using a much lower threshold. Taking this idea further, given a spatial definition of the target object predicted in the image, computational effort can be concentrated onto examining the image for merely the relevant features. These could be specified not only by expected feature groupings, but also by other attributes, for example texture and specularity. The less distinct the defining features of an object are, the more difficult it becomes to detect them by indiscriminate, data-driven methods, and the more benefit can be obtained by having a hypothesis-driven feature specification.

An Implementation

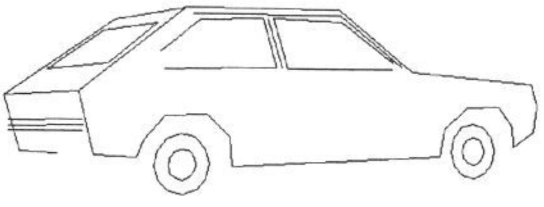
The model

The verification process starts with a detailed line-based three-dimensional geometric model of the target object (See figure 1), which uses facets to implement hidden-line removal. This provides an effective definition of the visual appearance of the object. Two-dimensional instances of this model when projected on to the image, give spatial information about the visible features of the object. (See figure 2). This model instance can then be matched against the data. In the present car models all the features are linear, thus simplifying the matching process.

AVC 1987 doi:10.5244/C.1.5



A Three-Dimensional Model of a Chevette
Figure 1



Two-Dimensional Instance of the Chevette Model
Figure 2

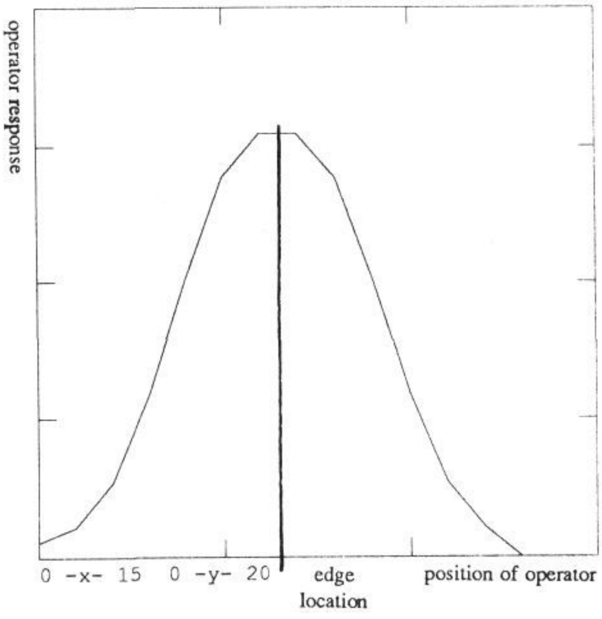
Evaluating features on the model

The projection of a 2-D instance invokes the evaluation of the features on the template. This evaluation is knowledge-based, i.e. the model provides details of the type and position of each feature predicted, thus only the relevant part of the image need be searched for the appropriate characteristic pattern. The evaluation takes place on data isomorphic to the image at the specified spatial scale. This has the advantage of keeping the model-matching time to a minimum, as multiple scale, image-wide databases of feature constructs and relationships do not have to be built and then searched through to discover the relevant feature in each case. The accuracy and robustness of the system is increased by matching directly to the image data. This makes relying on the correctness of the data-driven feature groupings unnecessary.

The actual method used for evaluation is based loosely on Canny's edge detector [Canny 1983]. He used maxima in the first derivative of a Gaussian of an image as evidence for edges, and maxima in the second derivative as evidence for bars. Both operators are directional; thus for smaller width detectors Canny took the derivatives in the x and y directions of a Gaussian blurred image and estimated the gradient direction from these partial derivatives. Directional non-maximum suppression was carried out along the gradient, to obtain local peaks of the derivative. Finally the edge elements were grouped by thresholding with hysteresis.

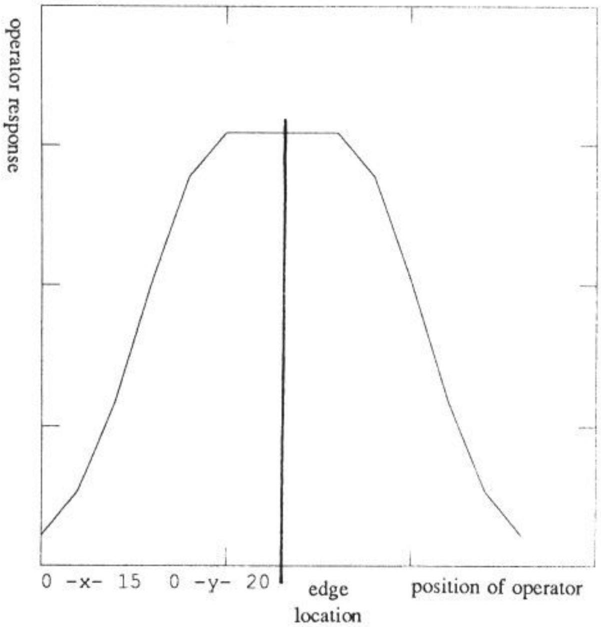
To implement a predictive operator a Gaussian filtered image was again used. Since the orientation of each feature is known, there is no need to estimate the gradient direction, but merely to take the derivative along the normals to the feature being tested. To evaluate an edge these normals are then searched for maxima and minima. For bars, instead of taking the second derivative, the signed slopes of the zero-crossings in the first derivative are used. (In the case of a symmmetrical bar this is mathematically equivalent). At each normal the absolute difference between the positive (maximum or positively sloped zero-crossing), and negative (minimum or negatively sloped zero-crossing) evidence is taken, and the average result along the length of the feature is taken to be the magnitude of the feature. Averaging the positive and negative result at each normal in this way makes the test selective for the predicted feature type; the bar test responds poorly to an ideal edge, and vice versa.

The size of the normal searched across in each feature test is linked to the size of the Gaussian used, and can also be varied to alter the specificity of the detector and allow it the ability to "cover" a range of misalignments in the line segment. (See Figures 3(a) & (b)). Gaussian filters with a standard deviation (sigma) of 1.5, 3 and 6 are used, together with normals of widths 1, 2, and 4 multiples of sigma to either side of the feature. This characteristic of the evaluator can also be used within a search strategy, so that a feature discovered by the operator with a range of several standard deviations of the Gaussian to either side of the line segment could be re-tested with the span of the normal reduced, in order to determine its position more accurately.



A graph of the output of the feature evaluator being applied at consecutive locations across an edge. The width of the normal is the same as the standard deviation of the filter being used
Figure 3(a)

A better method of pin-pointing the position of the feature is to do a line-fit to the feature evidence discovered at each normal. By fitting a cubic to the normal data, the position of the extremal point or zero-crossing can be determined to subpixel accuracy and the resulting series of points fed into a best-fit line algorithm. This approach has the added advantage that the deviation of the points from the best-fit line can be measured, and used to weight the result of the evaluator, to bias it against reporting a series of dispersed points as a feature. However in the present use of the evaluator, this has been omitted to improve speed.



A graph of the output of the feature evaluator applied to the same image data as in figure 3(a). Here the width of the normal is twice the standard deviation of the filter being used.
Figure 3(b)

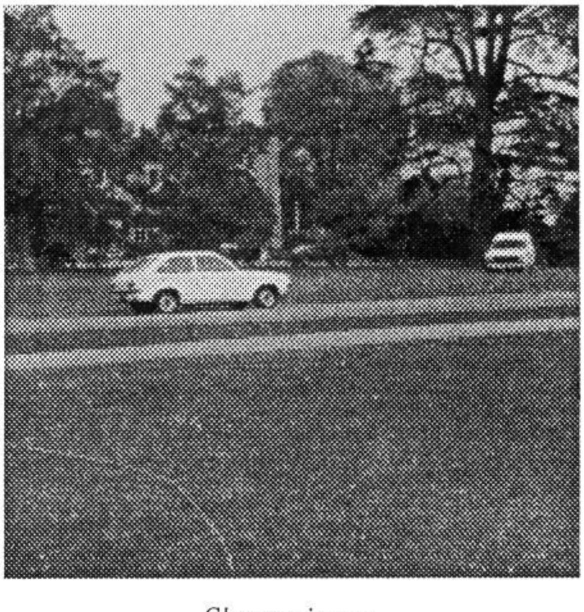
Combining the evaluations

The Chevette model used as an example here contains about 40 - 50 line segments, depending somewhat on the view. These line segments are evaluated separately and the individual scores are combined to produce an overall measure of the "goodness" of the match. The "magnitude" value returned by the evaluator is a relative measure of the strength of the feature being tested. It will depend on the circumstances of the evaluation: the standard deviation of the Gaussian filter used, the span of the normal, the kind of feature being evaluated, and of course the image being used, as well as the contrast of the feature. Therefore in order to be able to combine the separate scores meaningfully the result obtained in each case is weighted by the expected response of the evaluator to noise, in the same circumstances. This produces responses of greater than 1 for magnitudes better than that expected from merely noise, and values less or equal to 1 for magnitudes poorer or equal to the noise level. The geometric mean of the weighted responses was taken to be the score of the template.

The expected noise response of the evaluator to a particular feature is a measure of the average response to a randomly placed feature of the same type in that image. Testing the different features at a high density across an image would produce an estimate of this measure. However the variation of the noise response for each feature type with, the length, the Gaussian filter size, the search width and the standard deviation of the grey-level distribution in the filtered image, is fairly easy to characterise. It was expected that the noise response would be independent of length, since the measure we are taking is the average of the responses at each normal, but would be a function of the sigma of the Gaussian filter and the mean contrast of the image. In the case of edges the response should be independent of the width parameter, but for bars increasing the width allows broader bars to be recognised. Experiments were conducted which validated these expectations and determined values of the free parameters. Thus a set of simple relationships were derived for the different frequency gaussian filters used, relating the noise value for features assessed in that frequency to the contrast of the filtered image. The measure of contrast is the standard deviation of the Difference of Gaussian blurred image.

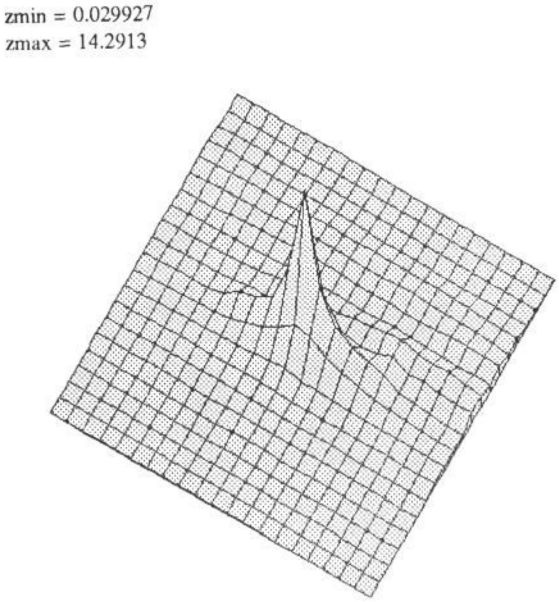
This method of combining the features in the image is an improvement over the system of setting arbitrary thresholds, in that the criteria for what is, and is not, significant is determined by the data in the image. At present the whole image is used in the noise analysis, but the system could be modified, so that the evaluator noise response is only determined for a portion of the image, i.e. within an area suggested by the low-level routines as being of interest.

Results

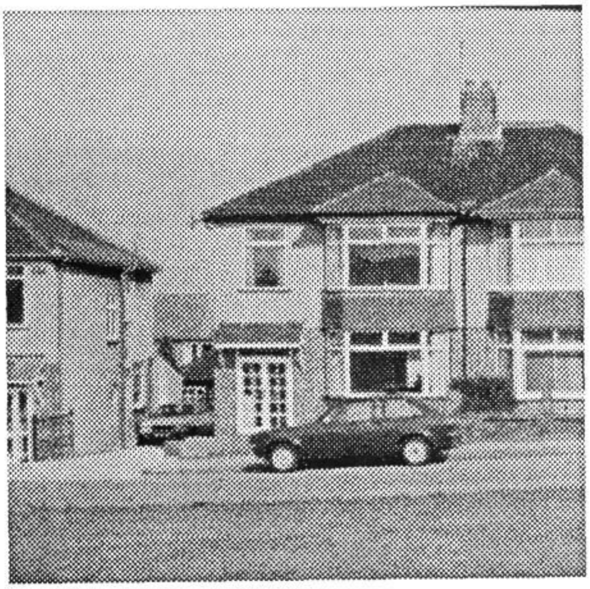


Chevette image
Figure 4

Extensive tests have been carried out on the evaluation procedures on 512x512 pixel images of natural, outdoor scenes containing vehicles, using a three-dimensional model of a Chevette as the target object. The system has shown a good discrimination between the target object and the irrelevant detail in the image. (See Figure 4 for an example image. Figure 2 is the correctly positioned 2-D instance for the car in figure 4. And figure 5 is a 3-D plot of the evaluation of the template placed at a grid of x,y locations)

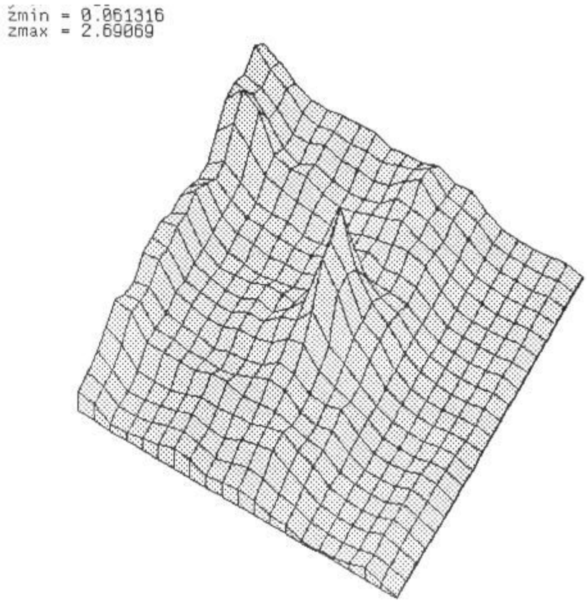


Example evaluation surface of a grid of y,z displacements of the 2-D Chevette template
Figure 5

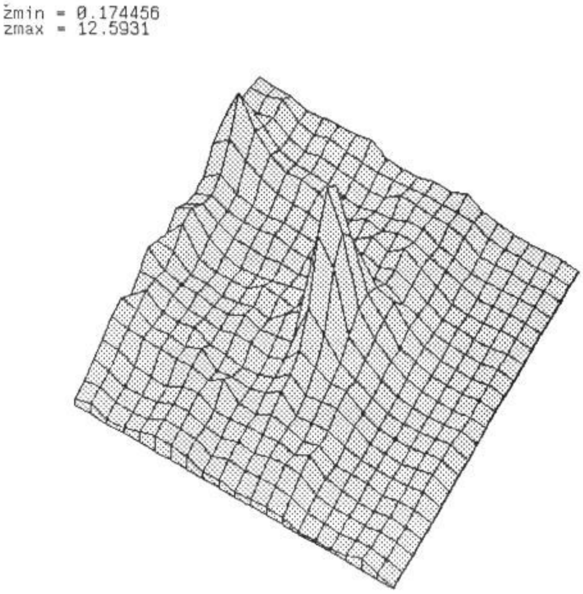


"Chevette in front of house" image
Figure 6

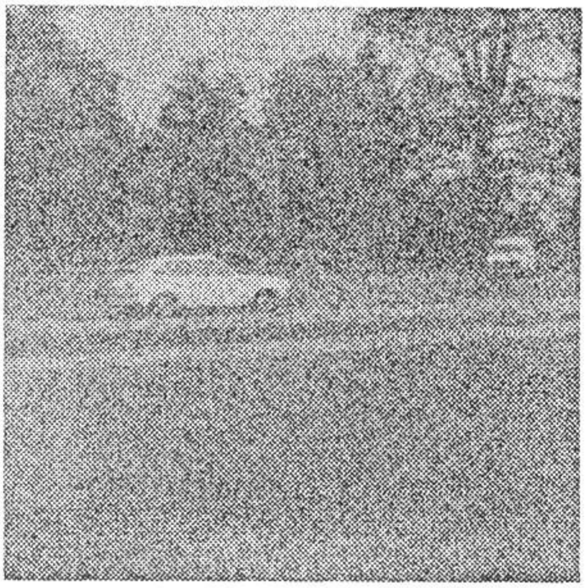
This technique has proved itself reasonably sensitive to errors in the positioning of the object for all six view parameters, It is not greatly effected by cluttered and noisy images, particularly if the sophistication of line-fitting is added to the evaluator. (See figures 6, 7 and 8 for an example of an image evaluated without and with line-fitting). Experiments are being conducted to test the performance of the evaluator on images with added white noise. (See figures 9 and 10 for an example of an image with added noise, and the resulting evaluation surface)



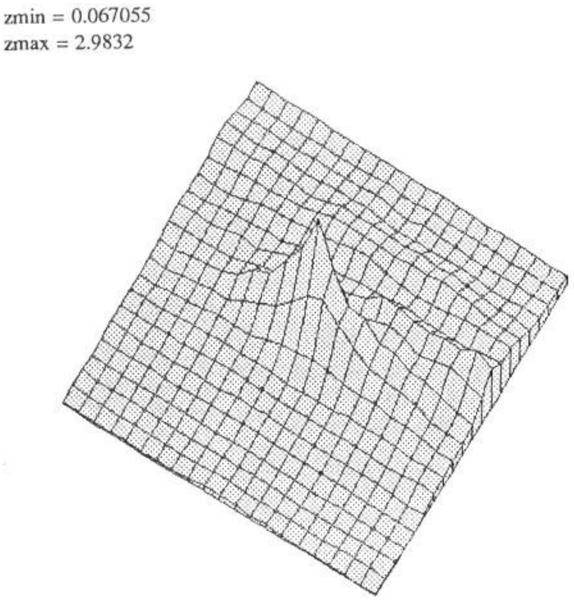
Example evaluation surface of a grid of x,y displacements of the 2-D Chevette template for the "Chevette in front of house" image
Figure 7



Example evaluation surface, as above, but with line-fitting added
Figure 8



The Chevette image in Figure 4 with reduced contrast and added Gaussian noise
Figure 9



Example evaluation surface of a grid of y,z displacements of the 2-D Chevette template for the noise-added image
Figure 10

Conclusions

Knowledge-based feature evaluation is a very effective means of object verification. The system used here has reliably detected hatchback cars in a variety of examples and has not been misled by non-examples. This has held true even in particularly noisy images. There are a number of improvements that could be added to the system. Firstly, not all the information encoded in the model description is used by the evaluator, for example the type of each feature, i.e. whether it is an extremum, fold, highlight, etc. There are also more constraints which could be included, for example forcing consistency of strength between features. At present support for the existence of a feature in the image is directly related to the absolute strength of the feature detected, rather than the correspondence of its strength value with that of the predicted feature. Another shortcoming is that at the moment only linear segments can be evaluated. Manufactured items tend to consist largely of straight-line segments, but the ability to evaluate curves would be useful. Work is also needed on the problems of occlusion and evaluating part models.

Acknowledgements

This work was carried out under the Alvey grant MMI-007. The modelling work was done by Anthony Worrall.

References

Brooks R.A. "Model-Based Computer Vision" UMI Research Press, 1984.

Canny J.F. "Finding Edges and Lines in Images" PhD thesis, MIT AI Lab, 1983.

Hutber D. and Sims P.F. "Use of Machine Learning to Generate Rules" in *AVC-87*, 1987.

Godden R.J., Fullwood J.A. and Hyde J. "Image Segmentation and Attribute Generation" in *AVC-87*, 1987.

Lowe D.G. "Three-Dimensional Object Recognition from Single Two-Dimensional Images" in *Artificial Intelligence* Vol 31, pp. 355-395, 1987.

Morton S.K. "Object Hypothesis by Evidential Reasoning" in *AVC-87*, 1987.

Shirai Y. "Recognition of Real-World Objects using Edge Cue" in *Computer Vision Systems* ed. A.R. Hanson & E.M. Riseman, Academic Press, 1978.