Alvey MMI-007 Vehicle Exemplar: Image Segmentation and Attribute Generation.

R.J.Godden, J.A.Fullwood and J.Hyde

Marconi Command and Control Systems Ltd., Technology Group, Camberley. GU16 5PE

Abstract

The MMI-007 project uses both bottom-up and top-down strategies to deal with the difficult problem of object recognition in outdoor scenes. Within this scheme, bottom-up methods fulfil the requirements of providing a concise image description suitable for symbolic reasoning and of providing an initial set of hypotheses to 'bootstrap' the top-down processes.

The nature of the imagery does not lend itself easily to the derivation of precise 3-D structural information by optical-flow or stereo techniques. This leads us to call upon a range of static segmentation techniques, each individually capable of capturing some aspect of the diffuse information present in our images. For example, surface, texture and colour homogenity, and boundary smoothness and continuity.

Introduction

The MMI-007 consortium's aim is to develop robust techniques for the analysis of unmanaged imagery, i.e. imagery arising from scenes for which there is little scope for arranging lighting or otherwise constraining the environment. The work so far has been concentrated on outdoor scenes; however similar problems could be expected in other unconstrained environments, for example, in the case of a mobile factory automaton.

The outdoor scenes we have worked with contain a diversity of object types at generally long ranges. There is a limited scope for the application direct methods, such as optical flow and stereo, of obtaining 3-D scene structure information. These methods could provide partial information however; for example, it may be possible to segment an optical flow field, even though no useful structure can be obtained from it. Motion is thus a potential source of segmentation information, although it has not yet been used within the MMI-007 exemplar.

Given that 3-D structure is not easily obtained we must make use of 2-D information as far as possible, leaving the higher level system to make inferences about relative depths etc., on an opportunistic basis, from other cues in the image.

The MMI-007 strategy is a mixed one, combining bottom-up (data driven) analysis, evidential reasoning and precise 3-D model-matching on specific objects of interest. The rationale behind this approach is discussed in Baker and Sullivan (1987) and (from the MCCS point-of-view) in Hyde *et al* (1985).

In the body of this paper we present some of the bottom-up techniques of relevance to the MMI-007 exemplar, along with some examples of the results. The techniques covered here come under two headings: Image Segmentation and Attribute Generation. Image segmentation in this context means those low level processes which operate on the input image to extract particular types of structural information, for example surface and boundary smoothness. Attribute generation covers the measurement of basic attributes from the segmentation primitives, through to the hypothesis of semantic labels, by hand-crafted or machine-induced rules.

Note : this paper purports to cover the work of the MMI-007 consortium in those areas within the exemplar context; however, the affiliation of the authors naturally leads to an emphasis on the work at MCCS. Where possible, references to the work at other sites are given for the reader wishing to gain a more balanced view of the consortium's work.

1. Image Segmentation

Because of the varied nature of the structural information in natural scenes there are many techniques which could be applied to yield part of the total 'picture'. The consortium has tended to concentrate on the following:

- Surface homogeneity, region-based approaches, appropriate to smooth (often man-made) objects.

AVC 1987 doi:10.5244/C.1.2

- Texture homogeneity, region-based approaches more appropriate for rough (mainly natural) objects.
- Boundary smoothness and continuity, edge-based approaches appropriate when fine profile detail or extended profiles are present.
- Colour, edge and region-based approaches making use of colour information.

Algorithms based on surface and texture homogeneity are presented below.

Edge-vs-Region based approaches

There is a tendency to regard these two approaches as competing alternatives, however within the context of a multi-algorithm approach there is scope for their complementary use. For example, region based surface-fitting techniques can segment smoothly changing image regions, producing closed-contour boundaries, where edge-based techniques may not succeed. Conversely, edge-based techniques are more appropriate for extracting information about lines and smooth curves. These are essentially independent processes which can provide support to one another. This is an interesting area which has not yet been taken up within the consortium, although others (Nazif and Levine, 1984) have looked at the problem.

Region Based Techniques

Two region-based techniques are described in some detail, firstly (section 2) a grey-level surface-fitting algorithm, which produces a multi-resolution region segmentation based on linear grey-level surfaces; second (section 3), an experiment in texture segmentation based on measures on the grey-level cooccurrence matrix (GLCM) and using an initial grey-level segmentation.

The surface-fitting algorithm is derived from an earlier region-merging algorithm, DSRM (Dynamic Smoothing, Region Merging) by Godden (1983). Recent work on DSRM has considered the effect of noise in the imagery and has resulted in a new region-merging criterion in which at each iteration a parameter S specifies a noise standard deviation, above which grey-level variations are considered significant. This approach has similarities with the work of Mowforth (1987), who in his ASL technique used an estimate of the image noise level to test the quality of fit of a surface computed over a local neighbourhood of each pixel; this surface is then used to compute a new grey-level for the centre pixel of each neighbourhood. ASL performs a noise-cleaning operation based on the known, or estimated, noise standard deviation. DSRM-SF (Surface-Fitting) achieves a similar effect, and also gives a segmentation into homogeneous surface regions. The sequence of resolutions output by DSRM-SF could be likened to several runs of ASL with increasing noise standard deviation.

Edge-Based Techniques

The VISIVE system developed at BAe implements an edge-based approach to segmentation. In recent months first-difference methods have been developed to replace the previous second-difference methods. Ths system is capable of accurate boundary placement (to sub-pixel accuracies) and gives some very good results on the MMI-007 exemplar data set.

First difference colour edge techniques have been used at STL, which in combination with intensity edges can give improved results.

2. Surface Fitting Segmentation

Algorithms for segmenting images into regions of near uniform grey level perform poorly when presented with images of objects whose surfaces display a smoothly varying intensity level. In many cases this variation arises from the surface being curved. The usual manifestation is the generation of a large number of regions aligned with the grey level contours. A segmentation of such images into regions which represent the surfaces as a whole is usually a preferable result. Such a 'surface fitting' operation has been incorporated into the DSRM segmentation algorithm.

The DSRM algorithms perform a region segmentation in an iterative manner so that the regions that result from the previous iteration are merged subject to the result of a test applied to the region's properties, referred to as the 'region merging criterion'. This criterion involves comparing the values of attributes (e.g. average grey level) of adjacent regions with a threshold. By gradually increasing the threshold from a low value at the first iteration (when the initial set of regions are the individual pixels), to higher values at successive iterations, segmentations corresponding to different 'resolutions' are obtained. These 'resolutions' are characterised by the significance of attribute differences, rather than by spatial properties.

In the DSRM-SF algorithm, the region attributes are the three parameters $\hat{g}_0$, $\hat{g}_x$ and $\hat{g}_y$. These parameters are calculated from the grey levels of the pixels within the region and characterise the first-order variation of grey level with image position in the region, so that the original grey level $g(i, j)$ of a pixel at position (i, j) relative to the region centroid can be approximated by

$$g(i, j) = \hat{g}_0 + i \cdot \hat{g}_x + j \cdot \hat{g}_y. \qquad (1)$$

Consequently, problems associated with estimating grey level gradients with fixed mask operators in the vicinity of region boundaries are avoided.

The criterion for merging a pair of adjacent regions is based on testing the hypothesis that in the presence of additive Gaussian noise, the observed grey levels in the two regions originate from just one popula-

tion. The regions are merged when the hypothesis is confirmed, and thus the merging of regions is, in general, more likely when the noise level ascribed to the iteration is large and when the regions are small in area. Furthermore, in the case of DSRM-SF, the merging criterion also depends on region shape as described later.

The noise level is defined by specifying a parameter S at each iteration, which is a measure of noise standard deviation. By increasing S from a low value (usually zero) at the first iteration, to higher values at subsequent iterations, segmentations corresponding to increasingly coarse resolutions are obtained. Here, resolution is characterised by the noise standard deviation S and can be interpreted as follows. At the first iteration, S is set to zero, and the algorithm regards all grey level structure in the image as 'genuine' and with no noise-induced component. At later iterations when the segmentations have been produced with a positive value of S, a certain amount of the structure in the imagery is regarded as induced by noise. The segmented regions correspond to clusters of pixels in which the underlying or genuine grey level structure (after the noise has been removed) is a linear function of image position. With large values of S even apparently genuine structure in the image is regarded as a noise effect and a coarse resolution segmentation results.

Description of the algorithm

In common with all DSRM algorithms, DSRM-SF performs a region segmentation in an iterative manner so that within each iteration the (old) set of regions existing at the start of the iteration is transformed into a new set of regions for input to the next iteration or for final output. The set of regions at the start of the first iteration is the set of individual pixels in the image. Each of the regions in the new set comprises at least one of the regions in the old set, and consequently all the pixels in those regions. The merging of regions is initiated at each iteration by taking the first region in the set of old regions as the seed for the first new region and successively including in the new region all old regions that are adjacent to it as it evolves and which satisfy the 'region merging criterion'. The particular form of this criterion for the surface fitting application will be described later. Subsequent new regions are formed in a similar manner, only including in the new region those old regions that are not already included in a new region at the current iteration. The result of the region merging is therefore dependent on the order in which regions are stored, but the effect of this is greatly reduced by making the region merging criterion most restrictive at the first iterations and successively relaxing it at later iterations.

The region merging criterion used in the DSRM-SF algorithm is best described by considering the situation after a few iterations have passed and some merging of regions has taken place. Before the start of the next iteration the grey level distribution of each region is described by three parameters, $\hat{g}_0$, $\hat{g}_x$ and $\hat{g}_y$ ($\hat{\underline{g}}$) which have been calculated by a least square method from the grey levels and coordinates of the pixels within the region so that the original grey level $g(i, j)$ of a pixel in a region at a position (i, j) with respect to the region centroid can be approximated by equation (1) in which $\hat{g}_0$, $\hat{g}_x$ and $\hat{g}_y$ are derived by first estimating the grey level gradient $\hat{g}_1$ and its direction $\hat{\theta}$ and then finding the two first order components and the zero order term by

$$\hat{g}_x = \hat{g}_1 \cos(\hat{\theta}), \quad \hat{g}_y = \hat{g}_1 sin(\hat{\theta}) \quad \text{and} \quad \hat{g}_0 = \frac{G_0}{N},$$

where

$$\hat{\theta} = \tan^{-1}\left[\frac{X_2 G_y - Z_2 G_x}{Y_2 G_x - Z_2 G_y}\right],$$

$$\hat{g}_1 = \frac{G_x}{X_2 \cos(\hat{\theta}) + Z_2 \sin(\hat{\theta})}$$

and

$$X_2 = \sum_R i^2, \qquad Y_2 = \sum_R j^2, \qquad Z_2 = \sum_R ij$$

$$G_0 = \sum_R g(i, j), \quad G_x = \sum_R i \cdot g(i, j), \quad G_y = \sum_R j \cdot g(i, j)$$

N = number of pixels in the region, and

$\sum_R$ is a summation over all the pixels in the region.

The parameters $\hat{g}_x$ and $\hat{g}_y$ are estimated by the above method because, in situations in which the region is one pixel wide, $\hat{g}_x$ and $\hat{g}_y$ could be both undefined if estimated by a more direct method. This method, by default, regards the component of the grey level gradient across the width of such regions to be zero, whence the computed components $\hat{g}_x$ and $\hat{g}_y$ derive only from $\hat{g}_1$ which is defined and oriented at an angle $\hat{\theta}$ along the length of the region.

It is convenient now to hypothesise that the true grey level parameterisation for the region is $\underline{h}=(h_0,h_x,h_y)$ and that the observed grey level $g(i, j)$ of the pixel at (i, j) with respect to the region centroid is due to independent noise components $n(i, j)$ added to the true grey level for the pixels, viz

$$g(i, j) = h_0 + i \cdot h_x + j \cdot h_y + n(i, j).$$

It can be shown that $\underline{\mathbf{h}}$ can be related to $\hat{\underline{\mathbf{g}}}$ as follows:

$$h_0 = \hat{g}_0 - \frac{N_0}{N}, \tag{2}$$

$$h_x = \hat{g}_x - \frac{Y_2 N_x - Z_2 N_y}{X_2 Y_2 - Z_2^2}, \tag{3}$$

$$h_y = \hat{g}_y - \frac{X_2 N_y - Z_2 N_x}{X_2 Y_2 - Z_2^2}, \tag{4}$$

where

$$N_0 = \sum_R n(i, j), \quad N_x = \sum_R i \cdot n(i, j) \quad \text{and} \quad N_y = \sum_R j \cdot n(i, j).$$

The values of N_0, N_x and N_y are not available from the imagery, but if the noise statistics are known then their statistical nature can be deduced. Assuming that the additive noise contribution of each pixel is governed by a Gaussian probability distribution function (pdf) with standard deviation S,

$$p(n) = \frac{1}{\sqrt{2\pi}\,S}\exp\left[\frac{-n^2}{2S^2}\right]$$

and then by making use of the characteristic function of this pdf as described in Davenport and Root (1958), the pdf of N_x is found to be

$$p(N_x) = \frac{1}{\sqrt{2\pi}\,S_x}\exp\left[\frac{-N_x^2}{2S_x^2}\right],$$

where

$$S_x^2 = S^2 X_2 .$$

Thus the value of N_x will lie in the range

$$-kS_x < N < kS_x$$

with probability erf(k). Taking a probability of 90%,

$$|N_x| < 1.64S\sqrt{X_2} . \qquad (5)$$

Similarly

$$|N_y| < 1.64S\sqrt{Y_2} \qquad (6)$$

and

$$|N_0| < 1.64S\sqrt{N} . \qquad (7)$$

Assuming that N_0, N_x and N_y are independent, which is true in all but the smallest regions and approximating the ellipsoidal constant probability contours of the joint pdf of N_0, N_x and N_y by a cuboid shape (described by equations (5 - 7)), it is possible to reorganise equations (2 - 4) and substitute for N_0, N_x and N_y giving

$$|N_0| = |N(h_0 - \hat{g}_0)| < 1.64S\sqrt{N} \qquad (8)$$

$$|N_x| = |X_2(h_x - \hat{g}_x) + Z_2(h_y - \hat{g}_y)| < 1.64S\sqrt{X_2}$$

$$|N_y| = |Y_2(h_y - \hat{g}_y) + Z_2(h_x - gx)| < 1.64S\sqrt{Y_2}$$

Each of these three equations can be regarded as a pair of parallel planes in the space of (h_0,h_x,h_y) between which exist possible values for the true parameter set $\underline{\mathbf{h}}$ for the region at the 90% confidence level and under the assumption that there is Gaussian noise of standard deviation S present. The space of possible values of $\underline{\mathbf{h}}$ forms a parallelepiped centred around $\hat{\mathbf{g}}$, the size of which is determined by S and the shape of which is determined by N, X_2 and Y_2.

On generating pairs of adjacent regions that are candidates for merging, the regions are allowed to merge if their associated parallelepipeds overlap. The volume that is common to both represents possible values of $\underline{\mathbf{h}}$ that could account for the observed grey level distributions in the two regions in the presence of noise of standard deviation S. However, account must be taken of the fact that the grey level parameterisation of each region is with respect to the centroids of the respective regions. It is therefore necessary to perform a transformation on $\hat{\mathbf{g}}$ at this stage so that all parameters are with respect to a single reference point. This only affects the zero order term. If the centroid of a region is (x_1,y_1) the effect of referencing with respect to (x_2,y_2) is to shift $\hat{g}_0$ to $\hat{g}_0'$, where

$$\hat{g}_0' = \hat{g}_0 + h_x(x_2 - x_1) + h_y(y_2 - y_1).$$

Then

$$|Nh_0 - h_x(x_2 - x_1) - h_y(y_2 - y_1) - \hat{g}_0| < 1.64S\sqrt{N}.$$

A further adjacent region may be merged if its parallelepiped intersects the common volume. The process continues, gradually reducing the size of the common volume until no more adjacent regions' parallelepipeds intersect with it, whereupon the next new region is initiated.

A second parameter to the algorithm, the threshold on the maximum grey level gradient H, is introduced by restricting the h_x and h_y axes of the feature space to lie between $\pm H$ and constructing parallelepipeds out of the space that remains.

The process continues until all regions have been processed at each iteration. A grey level image can be reconstructed after any iteration by performing the least squares estimation of $\hat{\mathbf{g}}$ for each region and using it to regenerate the noise-suppressed grey level for each pixel within the region.

DSRM-SF Results

In the results shown, the maximum grey-level gradient parameter H has been fixed at 25 grey-levels/pixel. The algorithm has run for a number of iterations on each example with parameter S gradually increasing from zero to a final value as given in the text.

Figure 1 shows a synthetic image which comprises two triangular regions, one having a horizontal linear gradient and the other one having a vertical linear gradient. Independent Gaussian noise with a standard deviation of two grey-levels has been added to this image. The image potentially poses a segmentation problem owing to the grey-levels on either side of the diagonal boundary being equal.

DSRM-SF was applied to this image both before and after adding the noise. When applied to the noise-free image the two regions are segmented precisely, as shown in figure 2 which depicts the region boundaries. The same result is obtained for final values of S between 1 and 300, when the two regions are merged into one. The image with added noise is also successfully segmented into two regions, as shown in figure 3, for values of S between 20 and 300. In this case there is some degradation of the boundary.

If we reconstruct an image from the surface coefficients of the regions ($\hat{g}_0$, $\hat{g}_x$, $\hat{g}_y$), the result shown

in figure 4 is obtained, demonstrating the noise cleaning ability of the algorithm. Figure 5 depicts the difference between the image reconstructed from the segmentation of the noisy image and the original noise-free image; the white areas have an absolute difference of less than two grey-levels, the black areas have a maximum absolute difference of six grey-levels.

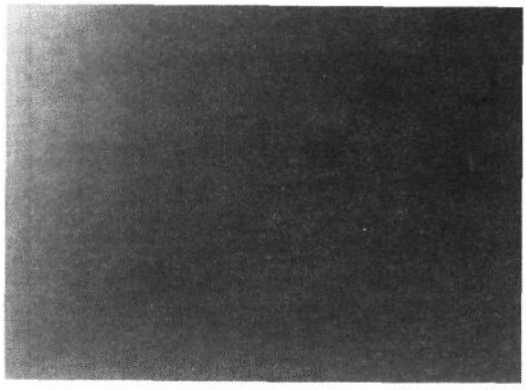

Figure 1 - *Image of Two Planar Regions*

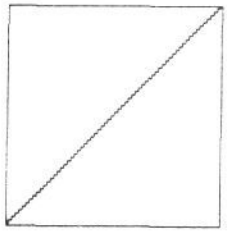

Figure 2 - *Segmentation of the No-Noise Image*

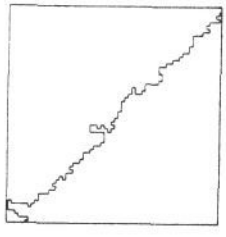

Figure 3 - *Segmentation of the Noisy Image*

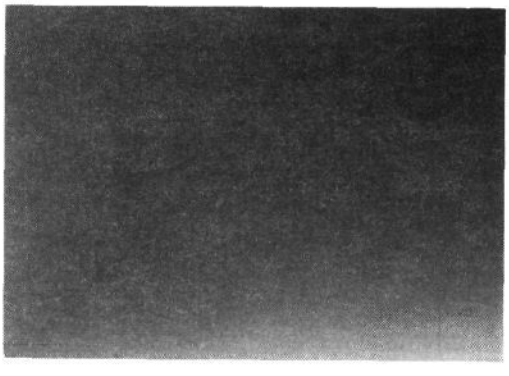

Figure 4 - *Grey-Level Reconstruction of the Noisy Image*

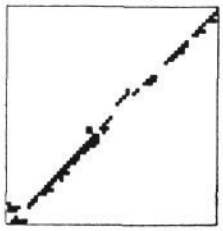

Figure 5 - *The Reconstruction Error, White is < 2 Grey-levels Error*

Figure 6 (overleaf) shows an image (RU01102) from the MMI-007 database. This was sampled at 128x128 pixels resolution before input to DSRM-SF. Figures 7 and 8 show two resolutions of output, for S = 20 and S = 65. In figure 7 detail such as the side-on car wheel hubs can be picked out, but large undulating areas such as the grass in the foreground contain unwanted boundaries. In figure 8, detail has been lost, but a much simpler description is obtained: in the foreground there are just a few major regions, and in the background large areas of trees have been merged. A grey-level reconstruction of the coarser resolution output (S = 65) is shown in figure 9.

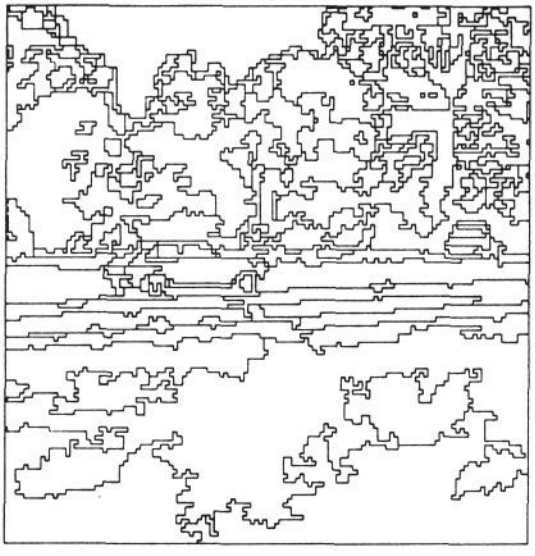

Figure 7 - *Segmentation of Figure 6, at S = 20*

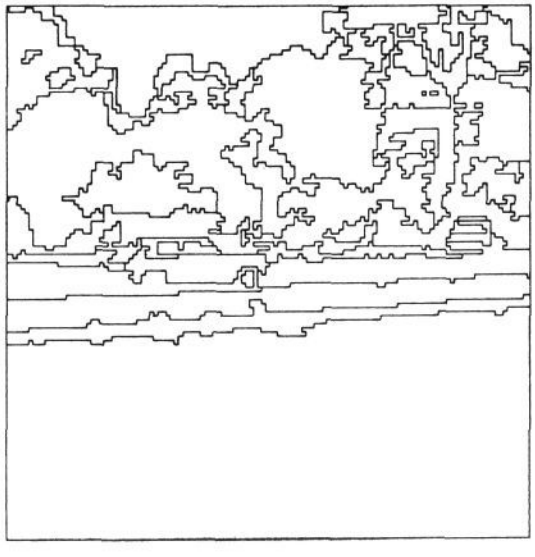

Figure 8 - *Segmentation of Figure 6, at S = 65*

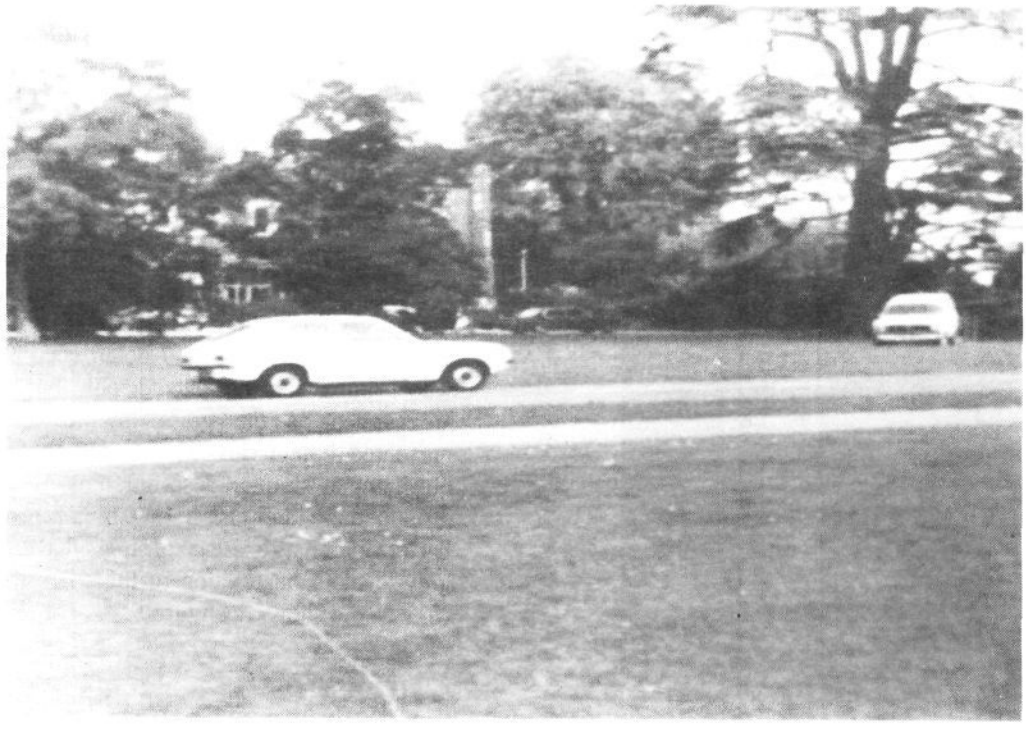

Figure 6 - *Database Image RU01102*

Figure 9 - *Grey-Level Reconstruction from the Segmentation of Figure 6*

DSRM-SF - Summary

The theoretical work described in this section forms the basis of a criterion for merging regions of linearly varying grey level which has been implemented in an algorithm DSRM-SF along the lines of DSRM. The evidence of the results obtained when DSRM-SF is applied to a set of synthetic and real images suggests that satisfactory results can be obtained.

It is thought that the region merging criterion may be usefully applied in alternative segmentation algorithms and that it can be naturally extended, if required, to handle higher order surfaces.

The current implementation of DSRM-SF on a VAX 11/780 segmented the two region (64x64 pixels) image, figure 1, in 12 seconds. More complex images such as figure 6, take about 60 seconds for 64x64 pixels. The time to produce several output resolutions is normally not much larger than the time to produce the first, as most of the time is spent early-on when there are many regions. Further work is required to explore approaches for efficient hardware implementation.

3. Texture Measures in Segmentation

The work described here uses measures derived from the grey level cooccurence matrix (GLCM) to achieve an image segmentation. The measures used are a subset of those known as the Haralick measures [Ballard&Brown, 82].

Previous workers have adopted the approach of dividing up the image regularly into regions, taking measures of texture over the regions and using them in a region-merging criterion. This approach is problematic because the use of regular initial regions does not take account of grey level and texture boundaries present in the image, resulting in a blurring of such boundaries and a consequent difficulty of segmentation. There is a trade-off between the large area needed for good texture estimates and the small area desired to minimise boundary effects.

The approach here is to use a primary segmentation of the image based on a grey level criterion only, until regions are obtained of sufficient size for meaningful texture estimates. The particular primary segmentation used here is the DSRM-SF algorithm, which produces a multi-resolution result in which the regions form a nested tree structure; but other techniques could be used instead.

Description of the Algorithm

The DSRM_SF algorithm operates iteratively to produce a series of region segmentations of decreasing resolution. Each region at a given resolution has a unique parent region in the next (lower) resolution; hence a tree structure of regions is formed, the extremes of which are the image pixels forming the leaves and the region encompassing the entire image forming the root.

For this work four resolutions of segmentation were used, with the coarsest resolution having between 10 and 50 regions, and subsequent resolutions having approximately double the number of regions of the previous resolution. Resolution 0 is defined as the root of the region tree - the region encompassing the entire image.

Segmenting the Image

Let m(k) be the number of regions at resolution k, k=0..n, m(0) = 1. We consider two resolutions, k and k+1, the parent regions are at resolution k and the child regions are at resolution k+1. For each parent region, texture measures are computed for each child region. If two child regions of the same parent have sufficiently similar texture measures they are merged and the tree structure modified.

Measurement and Comparison of Textures

The measures are derived from the GLCM, which is a joint-histogram of the occurence of grey level pairs for pixels having a given separation and orientation. Here, a separation of 1 and orientations of 0, 45, 90 and 135 degrees are used, giving four matrices, C(k), k=1..4. Three measures on the GLCM are used: Energy (E), Correlation (R) and Inertia (I).

If E_p,R_p,I_p and E_q,R_q,I_q are the measures over regions p and q respectively, both regions having the same parent, then we compute:

$$D = \min\left[\frac{\min(E_p,E_q)}{\max(E_p,E_q)}, \frac{\min(R_p,R_q)}{\max(R_p,R_q)}, \frac{\min(I_p,I_q)}{\max(I_p,I_q)}\right]$$

and, if D > P then regions p and q are merged. P is an input parameter ranging from 0 (always merge) to 1 (never merge).

Discussion of Results

The main effect of the algorithm is to reduce the number of regions in the tree areas, which as the main areas of significant texture in the MMI-007 database images. Thus in the image RU01107 (figure 10), the algorithm reduces the number of regions by about half, mostly in the tree areas. Figure 11 shows one resolution of a DSRM-SF segmentation input to the texture algorithm and figure 12 shows the output. In this example, parameter values were, for DSRM-SF: S = 33, and for the texture algorithm: P = 80. However, while gaining by the reduction in the number of regions in textured areas, on the side-on car the body and shadow regions have been merged; this is due to the lack of any significant texture in either area, at the resolution of the digitisation (512x512 pixels).

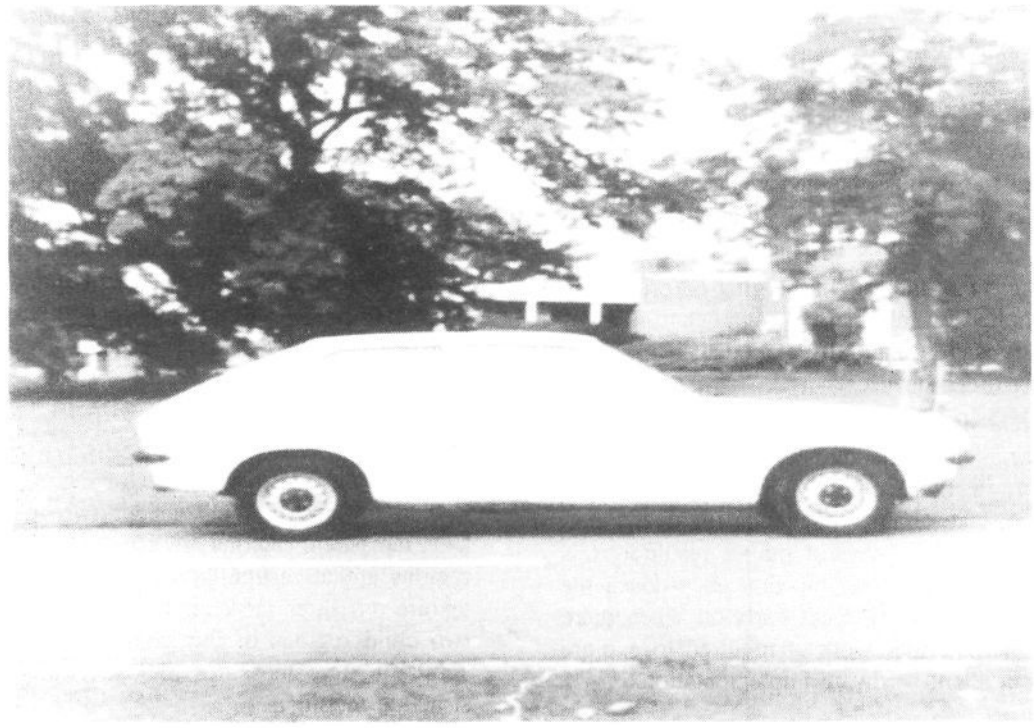

Figure 10 - *Database Image RU01107*

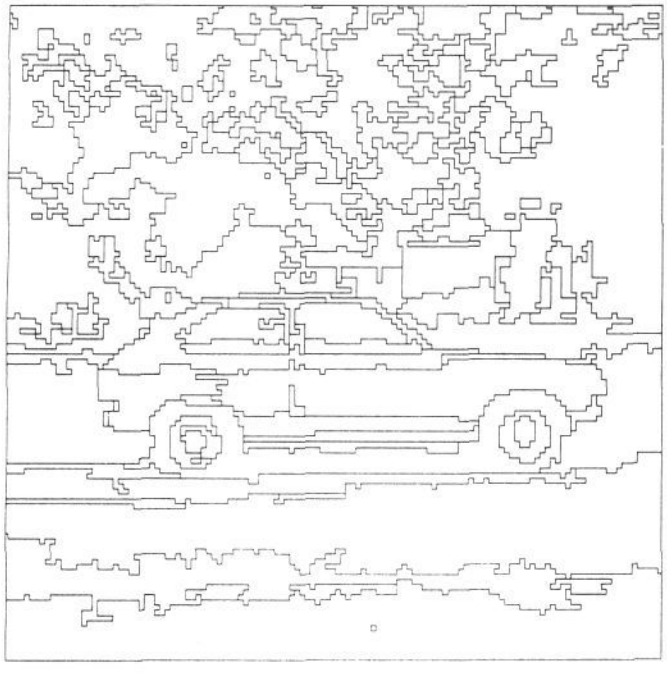

Figure 11 - *DSRM-SF Segmentation of Figure 10, S = 33*

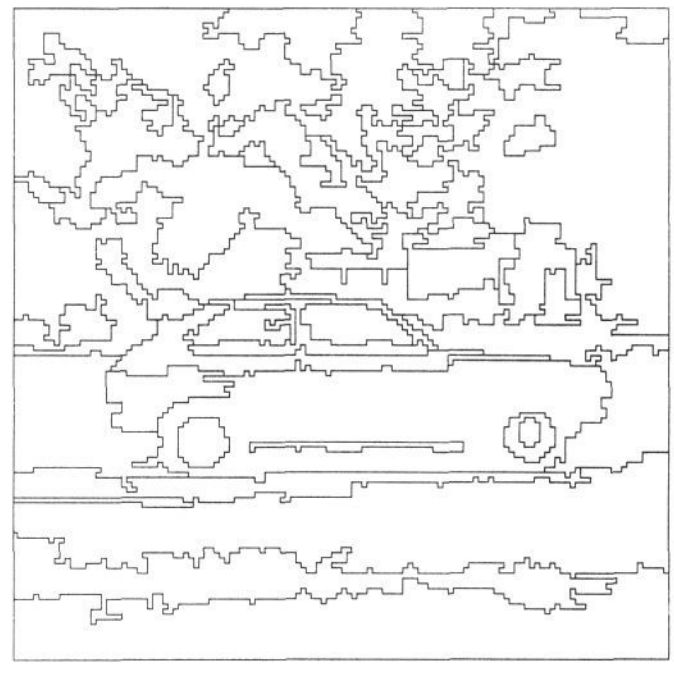

Figure 12 - *Texture Segmentation of Figure 10, P = 80*

4. Attribute Generation

The processes of evidental reasoning (Morton, 1987) and model matching (Brisdon, 1987) require inputs from the bottom-up processes; for the former as a factual description of the segmentation results in a suitable form; for the latter as hypotheses of good search areas, prioritised and constrained where possible. The rule-learning algorithm of Hutber and Sims (1987) also uses an attribute description of the segmented image, from which object recognition rules are derived.

The evidential reasoning process currently deals with region data derived from the DSRM-SF algorithm, hence it is in the form of a nested hierarchy or tree of regions, each level in the hierarchy forming one resolution of segmentation. The regions are each described by a set of features, most of which are straightforward measures over the image region - size, average colour, grey-level surface parameters, adjacency relations etc.

Less well-defined features are shape, of region boundaries and edge elements, and 'clues' - primitive hypotheses formed by low-level processes looking for characteristic features. Some examples are: the semicircular wheel-arches found in side views of vehicles and the parallel lines formed by the roof-windscreen-bumper lines in front-views.

We have implemented a 'wheel-arch' locator which searches for semicircular elements in the boundaries of a region segmentation. This finds both arches in the image RU01102 (figure 6) with one false alarm, see figure 13.

Figure 13 - *"Wheel Arches" Found in Figure 6*

We have also hand-coded rules which search for combinations of features, to give object hypotheses, for example: the side view of a car has a long region (body) with a smaller region above (window) and two small regions below (wheels). This complements the work of Hutber and Sims in automating the learning of vehicle classification rules: this provides an efficient method of finding patterns among feature vectors, which is not easy to do by hand; whereas hand-crafting provides a more abstract approach, which is very difficult to automate.

Summary

We have outlined two region-based segmentation techniques, based on surface and texture homogeneity. The two techniques are each appropriate to different object types, surface homogeneity for relatively smooth objects such as vehicles, and texture homogeneity for rough objects such as trees. Neither is entirely satisfactory at both tasks.

At present the two algorithms described, plus the edge and colour based techniques within the consortium, are separate processes. To make better use of each technique, a closer coupling between them is desireable.

We have mentioned some of the attributes derived from the image segmentations, which provide the basic descriptions used by the reasoning processes. These are also inputs to rule-based 'cuer' processes which hypothesise object instances; the rules being hand-crafted or machine induced. These, again, are currently separate processes which need to be more closely coupled in the future.

References

Baker, K. D. and Sullivan, G. D. (1987), The Knowledge Based Approach, AVC-87.

Ballard, D. H. and Brown, C. H. (1982), *Computer Vision*, Prentice-Hall.

Brisdon, K. (1987), Evaluation and Verification of Model Instances, AVC-87.

Davenport, W. B. and Root, W. L. (1958), *An Introduction To The Theory Of Random Signals And Noise*, McGraw Hill.

Godden, R. J. (1983), A New Dynamic Image Segmentation Algorithm Based on Region Merging, MCCS Working Paper #84.

Hutber, D. and Sims, P. F. (1987), Use of Machine Learning to Generate Rules, AVC-87

Hyde, J., Fullwood, J. A. and Corrall, D. R. (1985), An Approach to Knowledge- Driven Segmentation, IVC 3(4), Nov 1985.

Morton, S. (1987), Object Hypothesis By Evidential Reasoning, AVC-87.

Mowforth, P. H, Jelinek, J. and Jin, Z. P. (1987), An Appropriate Repesentation for Early Vision, PRLett 5(2), Feb 87.

Nazif, A. and Levine, M. (1984), Low-Level Image Segmentation - an Expert System, PAMI 6(5) Sept 84.