Image parsing for image retrieval from large image data bases: from coloured image to coloured regions

D. Sinclair
ORL,
24 Trumpington St.
Cambridge,
England.
email: *das@orl.co.uk*

Abstract

This paper gives details of a series of low level image processing routines which successfully break an image into a set of coloured regions. The first stage in the process is multi-scale edge detection. A fixed set of different sized kernels are used with the results being put into a single 'edgeness' image. A fixed lower threshold is applied to the edgeness image. A non-maximum suppression step is then applied. The histogram of edge strength in the non-max-suppressed image is used to set the high threshold for a hysteresis edge tracking routine. Points in the non-max-suppressed image above the high threshold are used as seed points to grow edges with the full edgeness image being used as the search domain. The edge growing algorithm therefore suffers less from the topological damage resulting from non-maximum suppression. A saliency filter is used to reject short crinkled edge chains.

The second stage in the process uses the edge image to generate a series of Voronoi peaks. These are used as seed points for dilation type region growing. As the regions are grown the edgeness image is used to give a simple integrated measure of distance from a Voronoi centre. Pixels are then assigned membership in the 'nearest' centre. Regions are not permitted to grow through edges in the edge image. A merging step is then applied to amalgamate regions with large non-edge shared boundary.

The mean colour of each region is evaluated and the regions then form nodes in a graph.

Keywords: Image retrieval, segmentation, multi-scale edge detection, region growing, colour regions, indexing.

1 Introduction

This paper provides a robust segmentation scheme for general coloured images. The segmented coloured region graph is intended to form the foundation of an image parsing suit for image retrieval from large image databases.

The first stage of the segmentation process is edge detection. Significant region boundaries are assumed to be correlated with edges in an image. Edge detection for image parsing must be concerned for with finding salient edges rather than with recovering every last detail and any thresholds required must either be fixed or estimated directly from the image.

A great deal of effort has been devoted to edges detection from the Marr-Hildreth [7] operator through filters designed to model the shape of an ideal edge [13, 1] and morphological attempts [8, 12] to phase congruency models [5] and multi-scale approaches [6, 3].

Edge detectors will typically employ some set of user defined thresholds to distinguish salient [9] from non-salient edges. Edge detectors that use a kernel or a set of kernels (modelling an ideal *edge* profile) to produce an *edgeness* image will typically have two edgeness thresholds; a low one which any pixel must exceed to be considered an edge and a high threshold which one pixel in a chain must exceed for all the pixels in the chain to be valid edgels. Edgel chain length is also widely used as a crude saliency measure: short chains are not important.

There are several criticisms that might be levelled against this family of approaches. Firstly they all rely on a **non-maximum suppression** step to thin the edgeness image and greatly reduce the number of pixels to be considered. This non-maximum suppression does a very poor job of preserving junctions with three or more branches. This is at least partly due to the fact that the kernels designed to model edges do not model junctions (with the exception of [12] where the local kernel shape is adaptive). Non-maximum suppression then destroys the local topology or connectivity of the edgeness image.

The next problem comes with the **edge following** routine that operates on the non-maximum suppress edgeness image. These edge following routines have either a four or an eight neighbourhood that they scan in a particular order when looking for the next pixel. This means that the edge follower will be biased in favour of left or right handed curves rather than straight lines. This means that an edge follower can be distracted by side chains rather than good straight edges. The follower also pays no head to the magnitude of the edgeness of the pixels it is following. Solutions to these problems have been proposed in [2, 4].

Lack of scale selection is also a problem with most of the current generation of model based edge detectors. Typically only one scale for the kernels modelling the ideal edge is employed and therefore only one scale of edge is detected. This weakness is hidden when a user is there to set the scale of the edge detector before a particular problem is tackled.

Explicit multi-scale edge detectors have been built, either using a fixed schedule of scales [6] or an adaptive one [10, 3]. They show considerable improvement in performance in images where the scale of the edges is unknown a priori or in images with a range of scales of edges.

The approach to edge detection taken in this paper is then a hybredisation of the best of the ideas from the above. A fixed schedule of different scaled derivative operators are used to produce a single edgeness image (implicit in this is a low threshold below which pixels are not considered as edgels). A non-maximum suppressed (NMS) edgeness image is produced from this. A histogram of the edgeness values is used to set a high threshold. Pixels in the NMS image above this value are are used as seed points for an edge tracking process. The edge tracking process uses the edgeness image as its domain of support (rather than the NMS image) and is biased in favour of straight lines.

Details of the algorithms and a visual comparison with a single scale Canny type edge detector are given in section 2

The next stage of the segmentation process is to use the edge map to generate regions. A Voronoi image is generated from the edge image, isolated peaks are used as seed points in a region growing routine. The edgeness image is used again in a distance measure to define the distance a point is from a Voronoi centre. Point are assigned memberships in their nearest centre. A merging step is applied to these regions and a graph with regions as nodes and adjacency relations as link is output. Details of the region growing technique are given in section 3.

Section 3.2 gives details of a shape descriptor designed to discriminate rectangular from non-rectangular shapes.

2 Edge detection

In this section details of the multi-scale edge tracker are given and a visual comparison with a generic single scale 'Canny' type edge detector is made.

The first stage in both edge detection schemes is to make an *edgeness* image from a grey-scale image (this is generated from the source colour image). Figure 1 shows an example colour image and its associated grey-scale version (it would of course be possible to perform edge detection on each of the colour planes but this is not done here).

The scale parameter in a generic 'Canny' type edge detector is then set as the σ values in a Gaussian smoothing kernel. The multi-scale edge tracker in this paper uses a fixed schedule of five different σ values (here $\{\sigma : (1.25, 1.5, 1.9, 2.5, 3.1)\}$). A single magnitude of gradient image is then composed as the maximum of the gradients of the five progressively smoothed versions of the image. In this case the gradient of each smoothed image is linearly scaled by σ following [6].

Figure 2 shows a multi-scale gradient image and two fixed scale gradient images. It is not generally known a priori at which scale salient edges will appear at in an image. Empirically it has proven satisfactory to choose a single low threshold to apply to all gradient images.

The next step is to perform non-maximum suppression on the gradient magnitude image. For the case of the standard 'Canny' type edge detector a hysteresis edge following routine is applied to the non-maximum suppressed image. Where hysteresis in this context means that on each continuous edgel chain there is at least one pixel with gradient magnitude above the high threshold and the rest above the low threshold. The high threshold is traditional set by hand.

For the case of the multi-scale edge detector the edge following routine functions

Figure 1: *Colour image of a golden retriever and his associated grey-scale portrait.*

Figure 2: *Magnitude of gradient images, multiscale* $\{\sigma : (1.25, 1.5, 1.9, 2.5, 3.1)\}$*, single scale* $\sigma 1.5$*, single scale* $\sigma 2.5$*. A low threshold was applied to the gradient images.*

differently. Pixels in the non-maximum suppressed image above a high threshold are used as seed points to grow edges. The high threshold is set automatically by making a histogram of the intensities in the NMS image and setting the high threshold so that 70 percent of edge pixels fall below it. The edge growing process functions as follows.

1. Find a seed point, a pixel in the NMS image above the adaptively derived high threshold.
2. For that point find the best grow direction. A 5 pixel long stick with 40 directions is used to sample the gradient magnitude image.
3. Move one pixel forward in the best grow direction.
4. If the growing process has a current direction then bias that direction when looking for the next pixel. Update the current grow direction.
5. Suppress seed points in the neighbourhood of the grown edge pixels.
6. Terminate if the value of the best grow direction falls below a threshold (this threshold is related to the high threshold).
7. Terminate if a growing edge hits another edge.

Figure 3 shows the effect of non-maximum suppression on the *edgeness* or gradient magnitude image in figure 2. This process destroys many junctions which are connected in the *edgeness* image.

The results of the edge tracker verses the traditional hysteresis edge follower are shown in figure 4. The edge tracker also has a build in saliency filter that rejects short crinkled edges. It should be noted that conventional edge following routines favour either left or right hand turns depending on the ordering of their local search domains. This can mean that they wander off the most salient edge and follow side chains.

The degree of straight biasing may be altered as may the *time constant* of the current direction estimator.

Further comparative results are given in figures 5 and 6. It can be seen that the proposed multi-scale edge growing algorithm produces edges with better connectivity and better saliency.

No current edge detector is perfect. The edge detector given in this paper was designed to operate on an arbitrary image without the benefit of an operator to select the 'right' scale and adjust the other internal parameters. It is also biased in favour of longer and hence more salient edges.

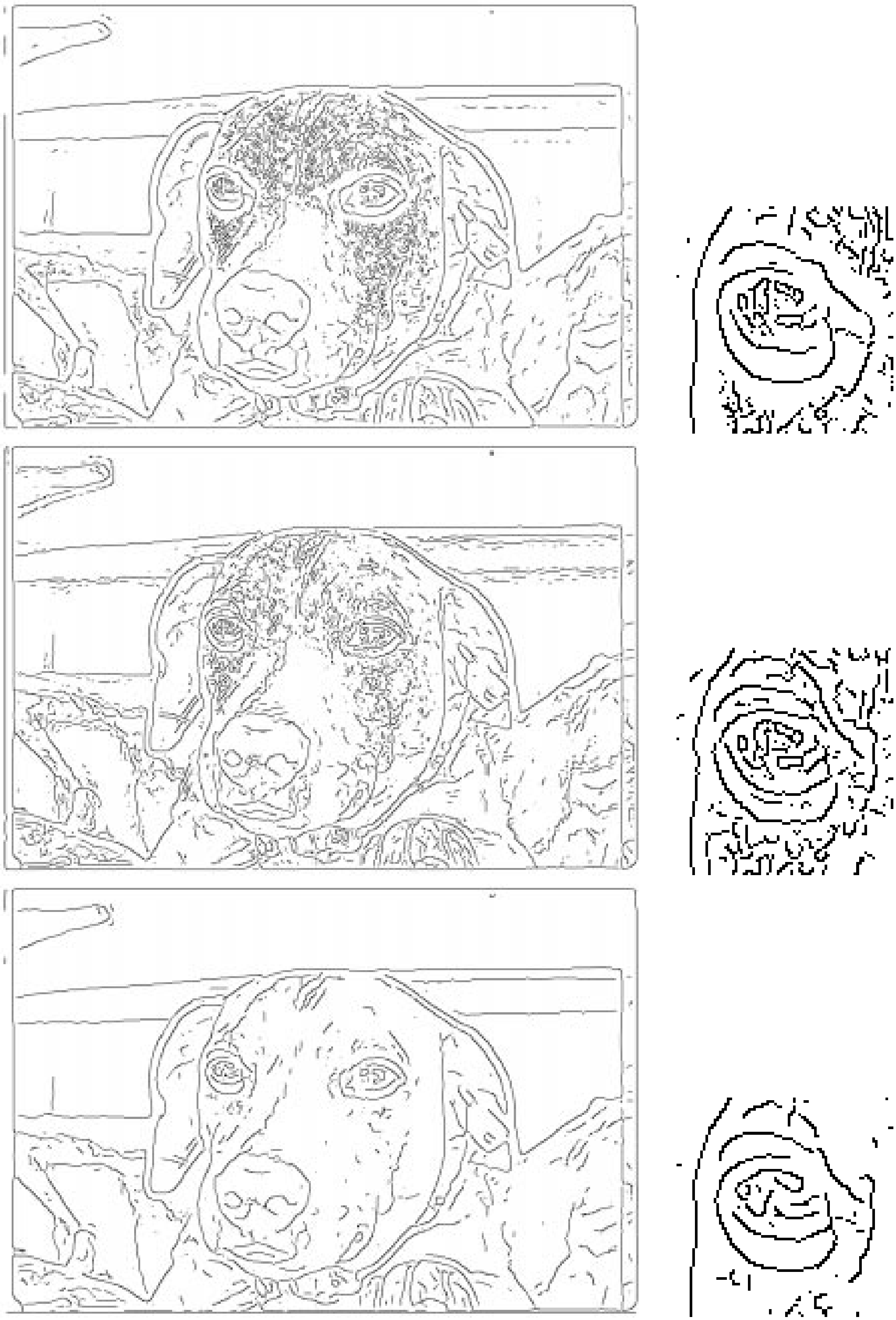

Figure 3: *Non-maximum suppressed versions on the images in figure 2. The images have been binarises for printing reasons.*

Figure 4: *Edge tracking results and hysteresis follower results applied to the respective images from figure 3.*

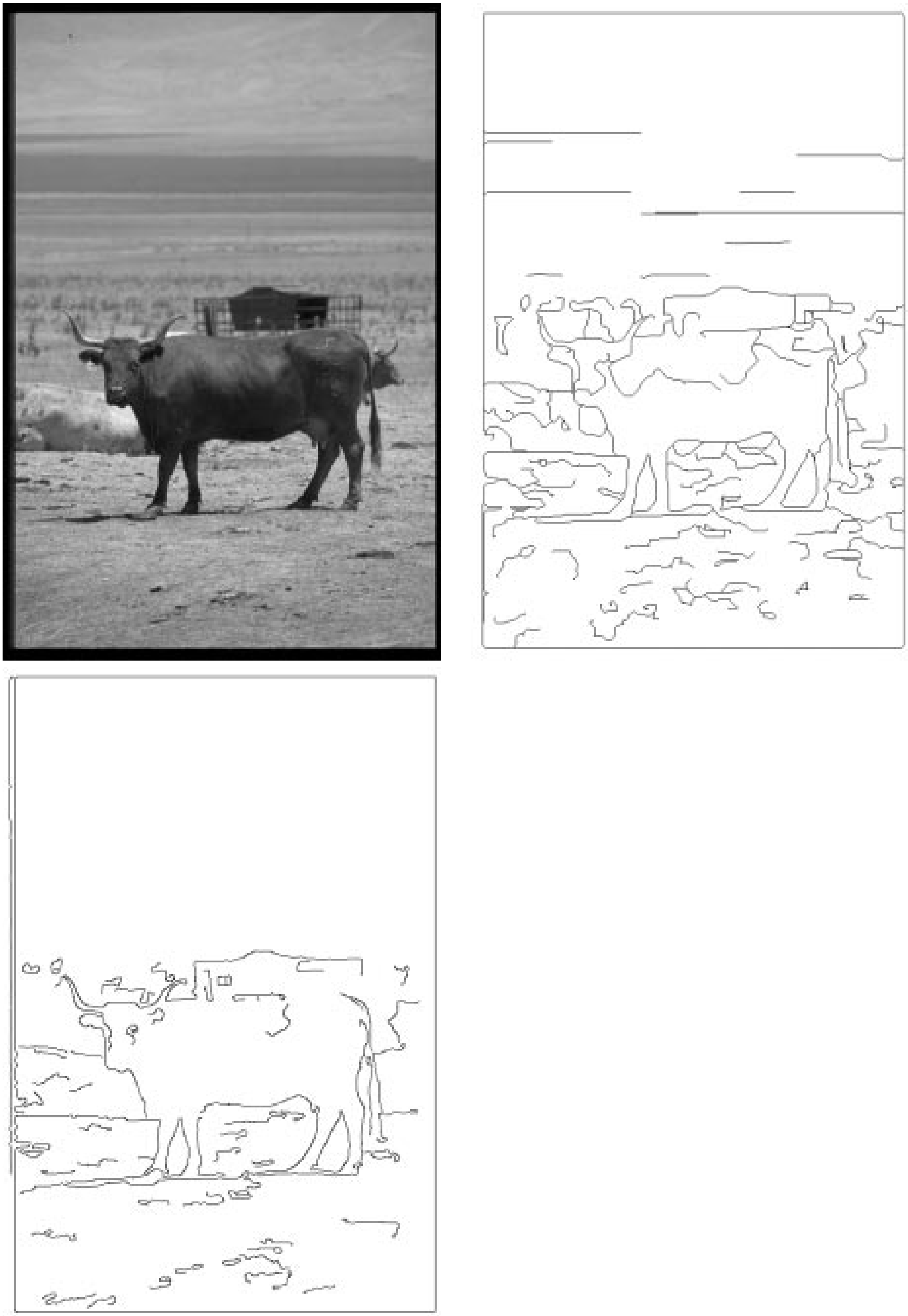

Figure 5: *Grey scale image of a cow, multi-scale edges, single scale* Canny *edges.*

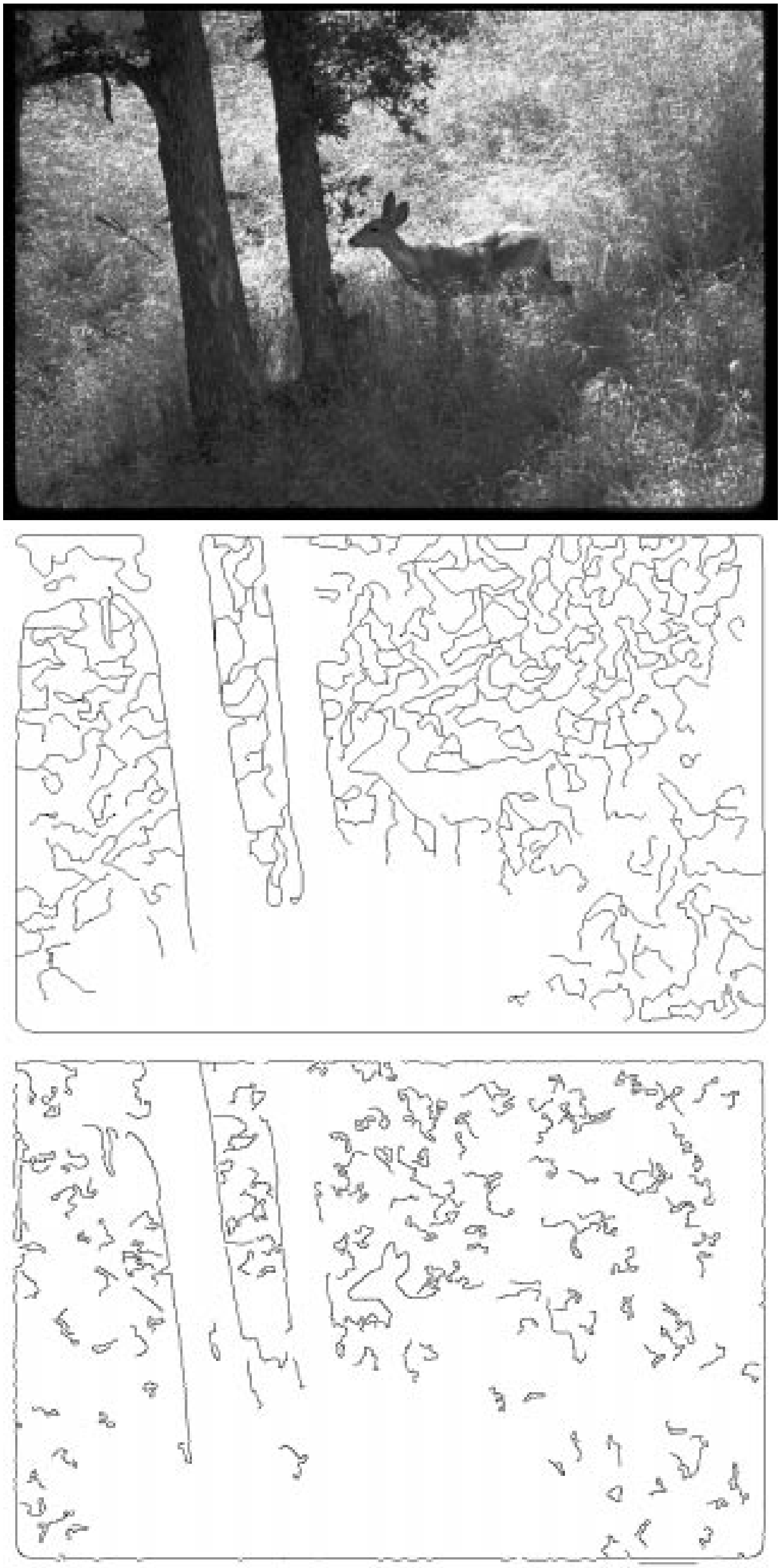

Figure 6: *Grey scale image of a dear, multi-scale edges, single scale* Canny *edges.*

3 Region growing

The first stage of the region growing algorithm is to compute the Voronoi image from the edge image (figure 7). In this the intensity at a pixel is the distance in pixels to the closest edge. In this case a 4 metric was used in the distance computation.

Figure 7: *Voronoi image computed from the edge image of the dog in figure 4. Because of a Manhattan metric for distances gives the image its characteristically square appearance. Peaks in the Voronoi image are used as seed points in the region growing routine.*

Peaks in the Voronoi image were then used as seed points for region growing. It was necessary to give peaks that were close together the same label. Regions were grown by propagation out from a region centre label. Each new pixel added to a region is assigned a distance from its label centre. Regions propagate outwards until they hit an edge or another region with a lower distance value to its label centre.

Regions are merged if their mutual boundary (direct region to region contact not contact across an image edge) form a significant percentage of one of the regions total boundary. The mean colour is computed for each of the new merged regions. Results are shown in figure 8.

Additional examples of the segmentation achieved are shown in figures 9 and 10.

3.1 Adjacency relations

The segmentation scheme makes use of adjacency relations for region merging. Adjacency information is also output as a graph with regions and their properties as nodes and adjacencies as links. Node properties currently include; area, mean colour, screen quadrant, first moment, squareness (defined in section 3.2), orientation and an adjacent region list.

Figure 8: *Coloured region segmentation achieved by the algorithm for the image in figure 1.*

3.2 Shape measures

Describing general shape remains a challenging and as yet unsolved problem in computer vision. Polygonal planar regions may be described by projective invariants [11]. These numbers have the advantage that they remain unchanged as the viewpoint from which the shape is seen changes. They do not however map readily onto a persons perception of shape.

The goal for a family of shape descriptors (for use in image retrieval) must be to allow a person to describe the shape of the object he wishes to find. Objects must be describable in terms of subparts with relations between subparts given in an intuitive way.

A first shape descriptor has been written but not yet evaluated. The descriptor measures *squareness*. The first and second moments of an image region are computed. These are used to define a rectangle with area equal to that of the of the region, aligned with the principle axis of the second moment, with aspect ratio the ratio of the square roots of the 2nd moments and centred on the 1st moment. Squareness is then defined to be the ratio of region pixels in the rectangle to the area of the region.

4 Discussion and conclusions

The edge detection method presented in this paper is of particular value for parsing large numbers of unfamiliar images because it gets over three of the main problems with generic *Canny* type edge detectors. The scale parameter does not need to be set by an operator as it is already multi-scale, the high hysteresis threshold is determined automatically and edge following is done in the gradient magnitude image rather than the non-maximum suppressed image resulting in improved topology for the recovered edges. The latter property is desirable for the region growing part of the motivating application of this paper. It is better to have too many regions rather than too few.

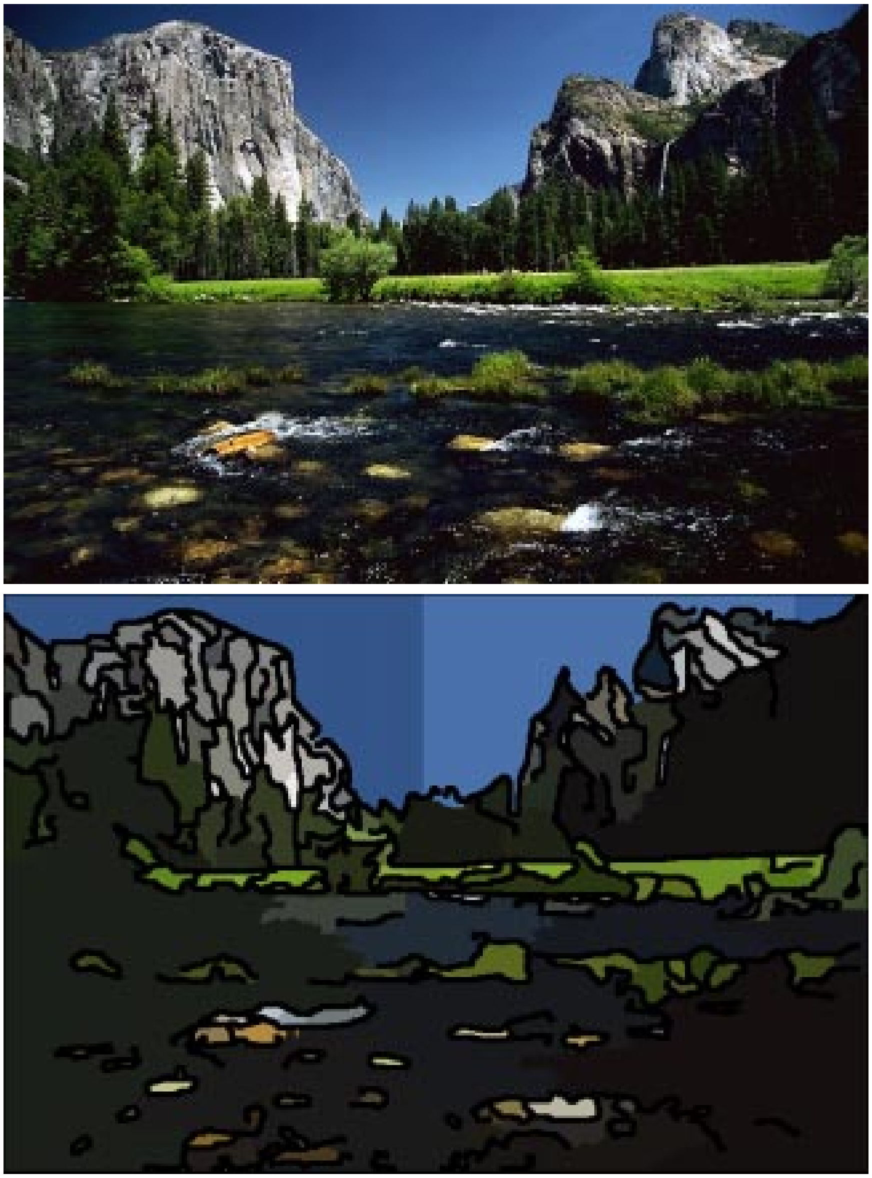

Figure 9: *Coloured image and associated colour region segmentation achieved by the algorithm.*

Figure 10: *Colour image of a man in a hat with a gun. Coloured region segmentation achieved by the algorithm.*

The edge detector clearly isn't perfect, comparison between edge map and edgeness image remains poor (but is better than straight Canny). The reason for this is perhaps perceptual edge saliency. The author is not sure how to address this problem further.

The segmentation presented is competent and is designed to form the basis for a full image parsing database search tool. Voronoi centres form reliable and intuitive seeds for the region growing process. Regions with a significant non-edge common boundary are merged. Visual inspection supports this step.

Regions based on grey scale edges are not guaranteed to contain homogeneous colour. Future version of the segmentation scheme will employ clustering to give a better description of the colour(s) present in a region.

It remains to perform an initial evaluation of colour, crude shape, adjacency and rough image location as indexing terms for image retrieval. It is thought likely that texture, special features (like faces and wheels etc.) and an improved shape description language will be desirable in a final system.

Issues of how to drive the focus of attention of an *interactive seed driven query* remain open and interesting problems.

5 Acknowledgements

Thanks to Prof Andie Hopper for steering this work closer to perception of regions and further from the combinatorics of blind edge interpretation.

References

[1] J.F. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intell.*, 8(6):679–698, 1986.

[2] S. Casadei and S. Mitter. A hierarchical approach to high resoluiton edge and corner detection. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 149–154, 1996.

[3] J. Edler and W. Zucker. Local scale control for edge detection and blur estimaiton. In *European Conference on Computer Vision*, pages 57–69, 1996.

[4] g. Hunt and R. Nelson. Lineal feature estraction by parallel stick growing. Technical note, SRI, 1994.

[5] P. Kovesi. Image features from phase congruency. Technical Report 95/4, University of Western Australia, Robotics and Vision Group, 1995.

[6] T. Lindeberg. Effective scale - a natural unit for measuring scale-space lifetime. *IEEE Transactions on Pattern Analysis and Machine Intelligence,*, 15:1068–1074, 1993.

[7] D. Marr and E. Hildreth. Theory of edge detection. *Proc. Roy. Soc. London. B.*, 207:187–217, 1980.

[8] J.A. Noble. Finding Corners. *Image and Vision Computing*, 6(2):121–128, May 1988.

[9] P. Rosin. Edges: Saliency measures and automatic thresholding. Technical note no. i.95.58, Ispra, 1995.

[10] C. Rothwell, J. Mundy, B. Hoffman, and Nguyen V.D. Driving vision by topology. Technical Report 2444, INRIA, 1994.

[11] C.A. Rothwell, A. Zisserman, D.A. Forsyth, and J.L. Mundy. *Fast Recognition using Algebraic Invariants*, volume Geometric Invariance in Computer Vision. Mundy, J.L. and Zisserman, A., editors, MIT Press, 1992.

[12] S. Smith. Feature based image understanding. D. phil thesis, University of Oxford, 1992.

[13] I. Sobel. On calibrating computer controlled cameras for for perceiving 3-d scenes. *Artificial Intelligence*, 5:185–198, 1974.