

Segmentation Coding Using Edge Detection and Region Merging

Yanbin Yu

R68/2.09 Rutherford Appleton Laboratory
Chilton, Didcot, Oxon, OX11 0QX, UK

ABSTRACT

An algorithm that integrates edge detection and region merging is presented. The algorithm can be employed as a pre-processing operation for model based image coding schemes. The pixels of the same intensities in an image are firstly clustered into small regions, resulting an over-segmented image where the boundaries are mainly edges. The further region merging operations are conducted in a way that the parameters are set to produce segmented image suitable for transform coding. Computer simulations he show that this algorithm yields suitable segmentation for effective transform coding.

1. INTRODUCTION

The applications of computer vision techniques to image coding appear great potential. Based on the model of the Human Vision System (HVS), the technical costs such as the number of bits to be transmitted over communications channels can be substantially reduced, while the fidelity is maximally retained. This new paradigm of image coding techniques, namely contour-texture or model-based methods, are regarded as a new generation of image coding algorithms and have attracted many researchers[1,2,3,4]. The basis of these methods is to preserve the crucial elements, such as edges and parametric boundaries of the regions, and degrade the image quality as *gracefully* as possible.

Moreover, the contour-texture coding algorithms can offer much more flexibility and adaptability than traditional redundancy-exploitation methods. A model based coding algorithm normally comprises two steps: segmentation and encoding. While traditional waveform-based algorithms becoming mature, it is not surprising that the techniques of computer vision like image segmentation are about to play an important role in image coding research.

Image segmentation is basically an *ad hoc* process. In general, a good segmentation method should be able to produce regions with homogeneous interiors and accurate boundaries. A plethora of segmentation methods has recently been developed[5]. Among them, the methods based operator such as the Canny[6] operator and the method based on zero-crossings of the second directional derivative[7] are rather popular. In such methods, edges can be well represented and segmentation operation can be conducted in a parallel manner. In Haralick's method, regions are initially formed by applying a connected components operation to non-edge

pixels. Then edges are assigned to their neighbourhood regions by region filling operations[7]. These schemes, however, do not take into account the homogeneous criteria in an explicit way. The number of regions is also beyond easy control. The split-merge method proposed by Klinger[8] is effective in the sense of making use of the homogeneity criterion[9], but fails to produce smooth boundaries. In general, operator based methods lead the results where edges are not necessary connected and somewhat over-produced, costing a huge number of bits for addressing. On contrast, region-growing methods[10] tend to produce false contours, presenting un-natural effects.

It is therefore necessary to develop a segmentation method, taking the advantages of these two techniques and avoid their respective shortcomings. This paper is intended to explore such a algorithm and is organised in following way: First, the requirements of final segmented results for image coding are studied. Then a two-step segmentation algorithm, integrating edge detection and region merging, is described. Finally, computer simulations and encoding results are presented.

2. THE REQUIREMENTS OF IMAGE SEGMENTATION FOR CODING

Firstly, it is known that human vernier acuity is highly sensitive to misalignment[11] of lines even to a few second of arc in visual angle. Edges, which define the shape of objects, are vital to the final perceptual result. In comparison with the other components, much more technical costs should be paid to preserve edges as faithfully as possible. In that zeroth-order approximation (or plateau) is applied, the interior of region should have the same amplitudes.

Secondly, extensive studies[12,13] have also proven that the DCT behaves very closely to the optimal transform in most image cases. This is particularly true in dealing with highly correlated image data. Ideally, the contents of the segmented regions should therefore be correlated and suitable for transform representation.

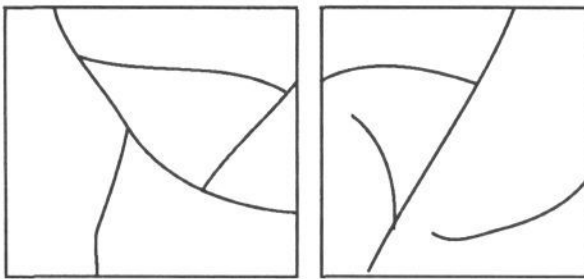
Thirdly, the number of segmented regions should be easily under control so that reasonable resources can be allocated to encoding of those boundaries.

With respect to image compression, the aim is to use the minimum number of functionals to approximate the data in a given region that minimises the error measure. The number of functionals should therefore be kept small to make compression possible. The error measure is normally MSE (Mean Square Error) or SSE (Sum Square Error) in general, and it should be stressed that such measures do not take into

account the characteristics of the HVS. However, in view of the inherent philosophy of the proposed algorithm, this error measure is still applicable.

3. THE SEGMENTATION ALGORITHM

The proposed methods are conducted in two steps: edge detection and region merging. The methods of performing edge detection are many. However, the first step is designed to produce the regions that enclosed with edges to satisfy the first requirement. It also ensures that the consequent operation is conducted on the basis of edges rather than artificial boundaries. The result of this step is an over-segmented image. Clearly, in order to maintain topological convenience, the edges should be all connected and regions are not allowed to be self-touched. That means any further eliminating of an edge will merge two adjacent regions.



(a) Valid Segmentation (b) Self-touched Graph

Figure 1. Topological Validation

Figure 1 is the examples of the edge map, in which (a) is valid in topology and (b) is a self-touched edge-graph. In the implementation, the Recursive Shortest Spanning Tree (RSST)[2], or Region Adjacent Graph (RAG) method[1] equivalently, can be used as edge detection and result is satisfied to the topological requirement. This method has a hierarchical structure and is guaranteed to produce connected regions. In brief, the algorithm iteratively joins adjacent regions based on some predefined criterion. The incoming image is mapped onto a graph, one pixel corresponding to one vertex. Costs for every possible merge are then calculated, sorted and stored in a table. In a situation where the image has 128 x 128 pixels and the graph is 4-connected, there are 2 x 128 x 127 initial connections, corresponding to the same number of possible mergings. Merging is always conducted in such a way that the distortion or cost caused by merging is minimised. It is therefore a stepwise optimal data driven approach, and there is no further restriction imposed on the shape of the final segmentation result. Moreover, it can also offer the possibility of embedding flexible merging criteria in context with application.

3.1 The Choice of Cost Function

The choice of cost function depends on the desired purpose of segmentation. A common cost measure is SSE (Sum Square Error):

where $g(x,y)$ is the approximation function and $h(x,y)$ is the original data over the given region. Biggar[2] has demonstrated that in the case of plateau approximation, when

$$sse = \sum_{x,y \in region} (g(x,y) - h(x,y))^2$$

$$g(x,y) = \frac{1}{n} \sum_{x,y \in region} h(x,y)$$

the algorithm will produce the maximum SNR with respect to the minimum number of regions. Also, the produced boundaries are all edges due to the inherent property of penalty calculation. In general, the criterion is the *maximum benefit return*[15].

Specifically, the criterion for merging is as follows:

$$C_{total} = C_{1+2} - C_1 - C_2$$

where C_{1+2} is the cost of coding the would-be merged region, C_1 and C_2 are the costs of coding regions 1 and 2, respectively.

In the case of transform coding, the following approximation functional is readily derived:

$$g(x,y) = T^{-1}QT[h(x,y)]$$

where $T[*]$ is the AST (Arbitrary Shape Transform)[14], and $T[*]^{-1}$ is the inverse AST and Q is the coefficient quantisation operation, respectively.

In general, one can either control the total bit-rate or limit the total degradation to achieve the coding requirements. Many criteria can be used to stop the final merging operation. A simple one, being the number of regions, relates to the bit-rate for encoding boundary information and should also be the same as the number of signal sources in the composite-source model. Likewise the alternative is to use the maximum SSE cost of merging two regions, controlling overall image quality.

However, two problems arise when the above cost calculation method is being applied to the segmentation.

One:

The boundaries of regions so produced will not correspond to edges. This is mainly because edges, referring to the sharp change in two pixel's amplitudes, can be approximated effectively by means of AST and that plateau representation cannot. As a result, in the early merging stages edges will melt into regions.

Two:

The number of operations involved is huge. For example, if image size is 128 x 128, to set up the initial cost table will

require 32,512 AST, quantisation and inverse AST operations. If a reasonable number of regions is required, another 30,000 operations are necessary to reach this segmentation. The overall computation is therefore prohibitively expensive.

Figure 2 and 3 show the segmentation results obtained by plateau approximation and AST coding.

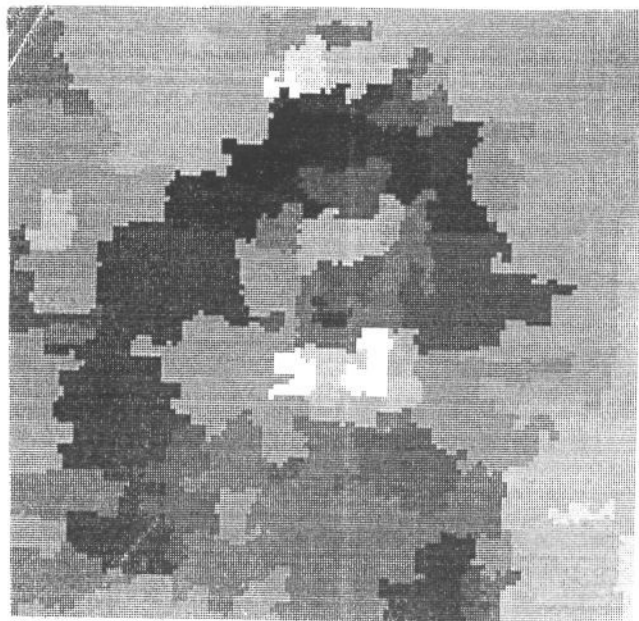


Figure 2. Segmentation by AST



Figure 3. Segmentation by Plateau

3.2. A Parameter N_p

A very important modification is made to solve the problem of edge melting into regions, while overall computation will also be reduced substantially. This modification is to maintain the ability of edge detection in the first place. The idea is based on following observations:

AST is capable of dealing with some rather sharp transitions, whereas the plateau approximation method is not. The cost calculation based on the plateau-approximation has edge detection properties, and requires much less computation as compared with AST method. In the meantime, adjacent pixels are very likely to have similar amplitudes, and should be merged in the first place.

Bearing the above considerations in mind, we can employ the plateau-approximation to set up the initial cost table and merge small regions until certain conditions are satisfied. After this initial merging, in which edges have been detected and preserved, and AST cost calculation is adopted. As a result, further merging operations will be guaranteed to produce regions encircled with edges. In pragmatic implementation, the stop conditions could be the number of merged regions by plateau approximation N_p . The optimal N_p depends on the content of the incoming image in an *ad hoc* way. In the simulation conducted in this research, it is found that the N_p should be 3 to 4 times the final required number of regions.

4. SEGMENTATION RESULTS

To assess the effectiveness of the proposed model, computer simulations have been conducted. The optimal number of segmented regions is, however, heavily dependent on the properties of the image. That is, if an incoming image consists of a few different signal sources, or data are highly correlated, a smaller number of segmented regions is preferred. On the contrary, if an image possesses many objects, or varying statistical characteristics in the spatial domain, a larger number of regions is preferred. Given the fact that this number will eventually influence the bit allocation for coding boundaries or edges, the number of regions should, therefore, be restricted to a reasonable range. By simulations, it was found that the optimal solution is likely to allocate 50 percent of the overall bit-rate to encode the boundaries and the rest to encode coefficients. For two typical 128 x 128 images, named *Virginia* and *George*, 150 regions are suitable if the overall bit rate is to be 1.0 to 1.5 bits/pixels.

One problem encountered after segmentation is the variation in region size. Since the segmentation is performed under the criterion of coefficient triangularisation[14], only the coefficients in a block of 8 x 8 residing at upper-left corner are retained. The zero-th order coefficients, or DC term, are scaled by a factor proportional to region area. For those very small regions, there are few pixels, so only the DC term is retained. For odd shaped regions, say, long and narrow in nature, again only the DC term is quantised. The remaining coefficients are quantised and encoded by Max-quantisation. Figure 4 and 5 present the reconstructed images of the proposed coding schemes at the bit rate of 1.0bit/pixel. Figure 6 and 7 are the respective segmentation images.



Figure 4. Reconstruction: Virginia
SNR = 19.13 dB



Figure 5. Reconstruction: George
SNR = 15.96 dB

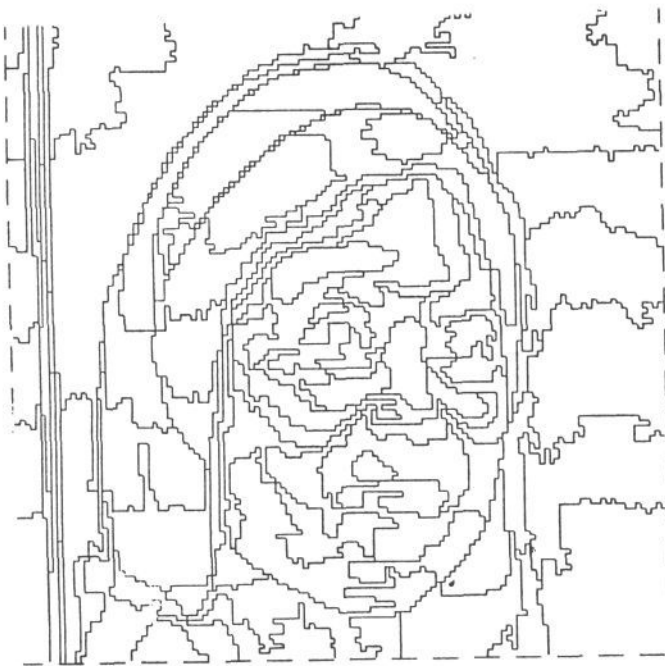


Figure 6. Segmentation: Virginia

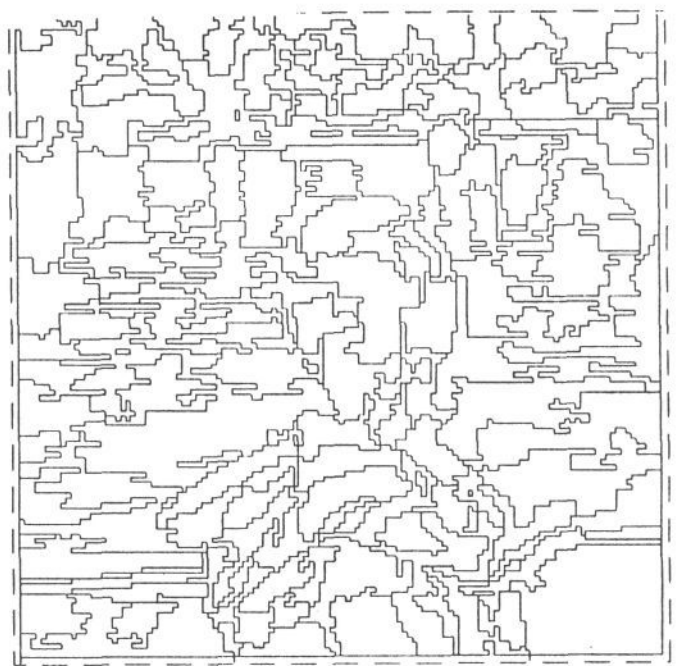


Figure 7. Segmentation: George

5. CONCLUSIONS

In this paper, an entirely new segmentation method is proposed and computer simulations are presented. The unique characteristic of the proposed algorithm is the integrating edge detection and region merging in context of transform image coding. The segmentation is based on the analysis of final requirements image coding and criterion of choosing parameters is also discussed. The computer simulation has shown that this segmentation method yields suitable segmentation for further transform coding operations.

ACKNOWLEDGEMENT

The author would like to thank Prof. A.G. Constantinides for his valuable suggestions and support.

REFERENCES

1. Kunt, M. Ikonomopoulos, and Kocher, M. "Second Generation Image Coding Techniques" *IEEE Proceeding*. Vol. 73 (1985) pp 549-574.
2. Biggar, M. *Source Coding of Segmented Digital Image*

and Video Signals. Ph.D. Thesis, Imperial College, University of London, London (1987)

3. **Gilge, M., Engelhardt, T. and Mehlan, R.** "Coding of Arbitrarily Shaped Image Segments Based on a Generalized Orthogonal Transform" *Image Communication, Signal Processing*, (1989)
4. **Yu, Y.B. and Constantinides, A.G.** "Pictorial Transform Coding for Tessellating Arbitrarily Shaped Regions" *To be published in Proceeding of V European Signal Processing Conference Barcelona* (1990)
5. **Haralick, R.M. and Shapiro, L.G.** "Image Segmentation Techniques" *Computer. Vision. Graphics and Image Process.* (1985) pp 100-132
6. **Canny, J.** "A Computational Approach to Edge Detection" *IEEE Trans. on PAMI*. Vol. PAMI-8, (1986) pp 679
7. **Haralicks, R.M.** "Digital Step Edges from Zero-Crossing of Second Directional Derivative" *IEEE Trans. on PAMI*, Vol. PAMI-8 (1986) pp 198-203.
8. **Klinger, A. and Dyer, C.R.** "Experiments on Picture Representation Using Regular Decomposition" *Computer. Vision. Graphics and Image Process.* Vol. 5 (1976) pp 68-105
9. **Kunt, M., Benard, M. and Leonardi, R.** "Recent Results in High-Compression Image Coding" *IEEE Trans. on Circuit and Systems*. Vol. CAS-34 (1987) pp 1306.
10. **Pavlidis, T. and Liow, Y.** "Integrating Region Growing and Edge Detection" *IEEE Trans. on PAMI*, Vol. PAMI-11 (1990) pp 225-233
11. **Foley-Fisher, J.A.** "The Effects of Target Line Length on Vernier Acuity in White and Blue Light" *Vision Research* Vol. 13 (1973) pp 1447-1454.
12. **Clarke, R.J.** *Transform Coding of Images* Academic Press, London (1985)
13. **Netravali, A.N. and Limb, J.O.** "Picture Coding: A Review" *Proceedings of IEEE*, Vol. 68, (1980) pp 366-406.
14. **Yu, Y.B.** *Transform Coding of Images Based on Shape Adaptive Models* Ph.D. Thesis, Imperial College, University of London, London (1990)

