

# EDGE POSTPROCESSING - A COMPARATIVE STUDY

E R Hancock\*<sup>†</sup> and J Kittler<sup>†</sup>

Rutherford Appleton Laboratory\*,  
Chilton, Didcot, Oxfordshire, OX11 0QX, UK

and

Department of Electronic and Electrical Engineering<sup>†</sup>,  
University of Surrey, Guildford, GU2 5XH, UK

---

*The use of contextual information is a vital ingredient in the successful postprocessing of edge information. Edge postprocessing is concerned with locating and labelling consistent edge-structures given confused or noisy response information. The available contextual information exists in the form of measurements provided by the gradient-vectors in pixel neighbourhoods and knowledge of acceptable edge structure which is expressed symbolically. We compare various approaches to the use of this information.*

---

Robust edge detection has been a goal of computer vision for almost three decades. Edge-detection algorithms reported in the literature are invariably multistage. Early stages are concerned with the characterisation of the raw edge information; this frequently involves filtering the raw pixel grey scale values to remove the effects of noise or produce a multiscale representation of the image data. Intermediate stages are concerned with the localisation of the edge response. Finally, the edges of physical objects are located by assigning labels to structures which exhibit desirable properties such as strong spatial connectivity. Canny [1] and Spacek [2] adhere to this multistage philosophy. Both authors derive filters for the optimal characterisation of step responses in the presence of noise. However, they adopt rather different methods for the postprocessing of the edge information. Canny performs non-maximum suppression by interpolating gradient vectors from an image neighbourhood and later draws upon connectivity information in the hysteresis linking stage. Spacek, on the other hand, imposes analytic continuity on the gradient magnitudes by fitting of a second-order polynomial surface to achieve non-maximum suppression without drawing directly on connectivity information.

The point that we wish to make in this paper is that although the characterisation of raw edge information has been addressed in a number of systematic ways, the postprocessing of this information has been treated less objectively. The postprocessing task is concerned with locating and labelling consistent edge-structures given confused or noisy response information. In practice it invariably draws on various sources of contextual information such as the measurements provided by the

gradient-vectors available in a neighbourhood of the image or the symbolic information conveyed by the connectivity of candidate edge structures. Although a number of schemes reported in the literature, including those described above, are certainly effective solutions to the problem, they do not objectively quantify the available contextual information. It is in this respect that contextual labelling schemes can be profitably applied to the edge-detection problem. Our aim in this paper is to stress the conceptual differences between the way in which edge postprocessing techniques draw on the available contextual information and the objective quantification of this information.

A host of contextual labelling strategies are described in the literature. The bulk of these are encompassed by the term relaxation. The common aim is to improve the consistency of object classification. This is achieved by utilising additional observational information in the form of measurements pertaining to objects and prior knowledge of the constraints which apply in a particular labelling application. The relaxation processes draw on this information in a number of distinct ways. For instance discrete relaxation is concerned with the updating of object-labels so as to achieve global consistency. Probabilistic relaxation, on the other hand, is concerned with updating label probabilities using the evidence provided by neighbouring objects. In a recent series of papers [3,4,5] we have detailed some extensions and improvements to the methodology of both types of relaxation process. An important element of the work has been the exploitation of the dictionary concept of Waltz [6] for the representation of labelling constraints.

We have drawn extensively on the postprocessing of edge information as an application vehicle to demonstrate the utility of the improved relaxation methodology [3,4,5]. These studies have shown that the dictionary provides a powerful way of representing the structure of the edge-process. Our aim in this paper is to compare the relaxation methods with the alternative edge postprocessing concepts reported in the literature. There are several aspects to this comparison. Firstly, we show how the different relaxation methods quantify and draw on the available contextual information. Secondly, we analyse how the same information is used by alternative post-processing algorithms. On the basis of this analysis, we select some algorithms for experimental comparison

with the relaxation approaches; this demonstrates some of the performance advantages of the relaxation methods.

## DICTIONARY-BASED RELAXATION

The task addressed by relaxation processes is that of assigning a symbolic label to each of an arrangement of objects through the exploitation of contextual information. In probabilistic relaxation the task is addressed as one of recursively combining evidence for the different competing labelling possibilities. In discrete relaxation, on the other hand, it is addressed by iteratively updating the symbolic label assignments to objects. In a recent series of papers [3,5] we have shown how the dictionary idea of Waltz can be used to represent the label process in improved formulations of discrete and probabilistic relaxation. The aim of this section is to review these improvements.

### Representation of labelling task

We are concerned with finding the class identities  $\theta_j$ ,  $\forall j \in \mathcal{N}$ , of the objects which have index set  $\mathcal{N}$  according to a label set  $\Omega$ . For the edge labelling application, the objects are pixels arranged on a regular square lattice. The label set used to represent the edge states encodes some limited directional information and is denoted by  $\Omega = \{\leftarrow, \uparrow, \rightarrow, \downarrow, \phi\}$ . The four arrows are orthogonal to the gradient directions of physical edges and  $\phi$  represents the possibility that a pixel belongs to the non-edge class.

The measurements available to the edge labelling task exists in the form of directional derivatives of the image in the directions of the pixel lattice axes. In both relaxation methods the measurements are characterised by the label probabilities for each pixel. In [4] we present a probabilistic model for the measurement process in the edge application. The model has two important features. Firstly, modelling of the measurement process is confined to considering the way in which noise manifests itself for non-edge pixels; it makes no assumptions of the appearance of edge features in the image luminance function and is therefore not restricted to step response. Secondly, the probability for the four edge labels is apportioned in a way that avoids bias in favour of edges propagating along the pixel lattice directions.

We denote the probability of label  $\omega_{\theta_j}$  on the object indexed  $j$  at iteration  $n$  of the relaxation process by  $P^n(\theta_j = \omega_{\theta_j})$  where  $\omega_{\theta_j} \in \Omega$ . In the probabilistic relaxation method these probabilities are updated in an evidence combining procedure. In discrete relaxation the label probabilities are static with iteration number; it is the label process that is updated to increase the probability of global labelling.

### The Dictionary Concept

Suppose that the object indexed  $j$  is directly interacting with the objects which have index-set  $I_j$ , where  $I_j \subset \mathcal{N}$ .

We are interested in constructing the dictionary of labellings over the objects in this index set. The permissible labellings for the neighbourhood  $I_j$  are listed in a dictionary denoted by  $\Theta_{I_j}$  which is of length  $Z(I_j)$ . Let  $\lambda_{\theta_l}^k$  denote the label on the object indexed  $l$ , corresponding to the  $k$ th entry in the dictionary  $\Theta_{I_j}$ , then we can introduce the following shorthand notation for entries in the dictionary

$$\Lambda_j^k = \{\theta_l = \lambda_{\theta_l}^k, \forall l \in I_j\}$$

with the dictionary denoted by the set of consistent labellings

$$\Theta_{I_j} = \{\Lambda_j^k, k = 1, Z(I_j)\} \quad (1)$$

We adopt the strict dictionary-model for the distribution of probability among label configurations. Accordingly, the probability measure associated with any physically impossible configuration  $\{\theta_l = \omega_{\theta_l}, \forall l \in I_j\} \notin \Theta_{I_j}$  is zero, ie

$$P(\{\theta_l = \omega_{\theta_l}, \forall l \in I_j\} \notin \Theta_{I_j}) = 0 \quad (2)$$

The available probability mass is distributed uniformly between the label configurations in the dictionary, ie they are assumed to be equiprobable

$$P(\{\theta_l = \omega_{\theta_l}, \forall l \in I_j\} \in \Theta_{I_j}) = \frac{1}{Z(I_j)} \quad (3)$$

For the edge application  $I_j$  is the 3x3 pixel neighbourhood. The constraints applying are that physical edges are continuous, a single pixel wide and undergo changes of direction infrequently. All labellings for the 3x3 neighbourhood satisfying these constraints have been compiled in a dictionary. There are 181 such labellings.

### Probabilistic Relaxation

Central to the original probabilistic relaxation approach suggested by Rosenfeld, Hummel and Zucker [7] is the iterative updating of the label probabilities for each object using an evidence combining formula. The updating of the label probability draws on contextual information through a support function for the object-label assignment, ie  $Q^{(n)}(\theta_j = \omega_{\theta_j})$ . The original evidence combining formula proposed by Rosenfeld, Hummel and Zucker, is

$$P^{(n+1)}(\theta_j = \omega_{\theta_j}) = \frac{P^{(n)}(\theta_j = \omega_{\theta_j})Q^{(n)}(\theta_j = \omega_{\theta_j})}{\sum_{\lambda \in \Omega} P^{(n)}(\theta_j = \lambda)Q^{(n)}(\theta_j = \lambda)} \quad (4)$$

One of our achievements in [3] has been the development of the following support function which draws on the dictionary of labellings for the entire context-conveying neighbourhood  $I_j$  and has an improved capacity to model the label process

$$Q^{(n)}(\theta_j = \omega_{\theta_j}) = \frac{1}{P(\theta_j = \omega_{\theta_j})} \cdot \frac{1}{Z(I_j)}$$

$$\sum_{\Lambda_j^k \in \Theta_{I_j} | \lambda_{\theta_j}^k = \omega_{\theta_j}} \left\{ \prod_{l \in I_j} \frac{P^{(n)}(\theta_l = \lambda_{\theta_l}^k)}{P(\theta_l = \lambda_{\theta_l}^k)} \right\} \quad (5)$$

Since at each update the current probabilities reflect contextual evidence from increasingly large neighbourhoods, the repeated updating would eventually achieve the desired objective of drawing global contextual information from the complete network.

### Discrete Relaxation

Our discrete relaxation procedure treats the label updating procedure as global MAP estimation. It seeks the configuration that maximises the quantity  $P(\theta_i = \omega_{\theta_i}, \forall i \in \mathcal{N} | \underline{x}, \forall i \in \mathcal{N})$  by iteratively updating single labels. There is a serious obstacle to modelling the joint probability of labels  $P(\theta_i = \omega_{\theta_i}, \forall i \in \mathcal{N})$  using the dictionary concept. If the initial label assignments are noisy or inconsistent, it may be impossible to create dictionary configurations by single label replacements. As a consequence it is difficult to determine the direction of optimal label update which increases the global consistency; the discrete relaxation procedure becomes deadlocked.

We have overcome the deadlock problem by introducing a new conceptual ingredient: We regard inconsistent labellings as corrupted dictionary items. The development of this idea requires a physical model of the label corrupting process. We have adopted a particularly simple model in which label errors are independent and occur with class-independent label probability  $P_e$ . According to the model the number of label errors on each neighbourhood system follows a binomial distribution. This has the attractive feature that the likelihood of observed label configurations is determined by their congruency with dictionary items. Following Hummel and Zucker [8], the label process measures the effect of label replacements on all neighbourhood systems containing the object of interest. Consider the effect of label replacement  $\theta_j = \omega$  at the object indexed  $j$ . The congruency measure between the dictionary item indexed  $k$  and the newly constructed label configuration on the neighbourhood system  $I_r$  (where  $j \in I_r$ ) is

$$K(r, j, k) = \epsilon(\omega, \lambda_{\theta_j}^k) + \sum_{l \in I_r, l \neq j} \epsilon(\omega_{\theta_l}, \lambda_{\theta_l}^k) \quad (6)$$

where  $\epsilon(i, j) = 0$  if  $i = j$  and  $\epsilon(i, j) = 1$  if  $i \neq j$ . Under the memoryless label corruption assumption the label process is modelled by the following criterion function

$$F(\theta_j = \omega) = \sum_{r \in I_j} \frac{b_r}{Z(I_r)} \sum_{\Lambda_r^k \in \Theta_{I_r}} a^{K(r, j, k)} \quad (7)$$

where  $a = \frac{P_e}{(1-P_e)}$  and  $b_r = (1 - P_e)^{|I_r|}$ . With these ingredients, the label update procedure is stated as a decision rule which assigns  $\theta_j$  to class  $\omega_{\theta_j}$  if

$$\frac{P^{(0)}(\theta_j = \omega_{\theta_j})}{P(\theta_j = \omega_{\theta_j})} F(\theta_j = \omega_{\theta_j})$$

$$= \max_{\omega \in \Omega} \frac{P^{(0)}(\theta_j = \omega)}{P(\theta_j = \omega)} F(\theta_j = \omega) \quad (8)$$

## THE ROLE OF CONTEXT

In this section we discuss how different edge postprocessing methods draw on the available edge information. We contrast techniques which aim at objectively quantising contextual information, such as dictionary-based relaxation and Markov models, with heuristic approaches. An important aspect of this comparison is an analysis of how the schemes draw on observational information and knowledge pertaining to edge structure.

### Dictionary-based relaxation

According to probabilistic relaxation, raw gradient information is regarded as evidence that should be combined. The information used in this process are the label probabilities in a 3x3 neighbourhood and the dictionary of consistent labellings. The process proceeds recursively until eventually the probabilities converge on a hard labelling; this is taken as the final interpretation of the edge information. The updating of label probabilities can be interpreted as an implicit filtering of observations.

Discrete relaxation regards the label probability for the object under consideration as one component of the evidence for a label assignment; the second component reflects the consistency of surrounding label assignments. By contrast with the probabilistic relaxation method, the label probabilities remain static and the dictionary is used to gauge the consistency of the assigned label configurations; this draws on the label assignments to the nine 3x3 neighbourhoods that form the 5x5 pixel window. In addition to the label and measurement processes, the method incorporates a label error process which is assumed to be memoryless. The label error probability may be used to control the discrete relaxation procedure; by incrementally reducing  $P_e$  inconsistent labellings are increasingly penalised. Contextual information is propagated not by combining evidence in object neighbourhoods but by updating the label process.

### Other Relaxation Approaches

The model described above should be contrasted with those adopted by Berthod and Faugeras [11] and by Zucker et al [12] in their applications of probabilistic relaxation to edge-labelling. According to both of these approaches, the initial label probabilities are assigned on the basis of the responses to a set of directional difference masks; no attempt is made to justify the way in which the raw measurements originate from realistic

edge or noise processes. A second point of contrast is the limited use of prior knowledge in the form of probabilities for labels on pixel pairs. In the work of Zucker [12], these probabilities took the form of heuristic compatibility coefficients in the range -1.0 to 1.0; these coefficients were chosen to encourage continuous edges and suppress corners. A similar goal orientated approach was adopted by Berthod and Faugeras [11] who derived a set of probabilities with the objective of encouraging continuous straight edges. Our method exploits more information concerning the structure of the edge-labelling application and draws on it in a less biased way.

## Markov Models

Geman and Geman [13] have developed a stochastic relaxation method that seeks the MAP labelling using a simulated annealing technique. The method has been applied to image segmentation. A Markov field is used to model boundary and region components of the label process. This model draws on an index of connectivity. However, since the Markov model admits any occurring boundary configuration with finite probability it is not faithful to the dictionary idea; this restricts the capacity of the label process to model consistency. It is difficult to assess independently the effectiveness of the boundary and region processes.

An interesting edge postprocessing technique has been reported by Haralick [14]. The method draws on a facet model characterisation of edge response and uses a Markov chain concept to model the label process. The Markov chain allows the label probabilities to be computed recursively by distinguishing between the causal past and future of continuous edge structures. The label process is defined on Markov chain and draws on pairwise relationships between adjacent edge labels; an independent decision is required to assign the edge direction. There is no dictionary concept underlying the approach; connectivity is encouraged by the chain indexing of pixels.

## Non-maximum Suppression

The idea of edge localisation through the suppression of non-maximum response was introduced by Rosenfeld and Thurston [16]. Given the edge magnitude and direction at each pixel, the underlying concept is to draw on additional information from the surrounding 3x3 pixel neighbourhood for the existence of localised edge structure. This is achieved by examining the gradient magnitudes for pixels perpendicular to the edge direction for the pixel under consideration. If these magnitudes are consistent with a local maximum of the gradient, then the edge hypothesis is retained. The method therefore draws on a model for idealised edge magnitude profiles in the measurement domain of gradient magnitudes; the information used consists on one edge direction and the edge magnitudes for three selected colinear pixels. There have been numerous extensions and refinements

of this basic idea. For instance Canny [1] uses a method that interpolates the gradient magnitudes in a 3x3 pixel window to determine the gradient-vectors of the local maxima.

## Surface models

The literature reports several approaches to edge detection that draw on surface models of the image luminance function. Nalwa [9] locates edges by fitting profiles which are motivated by considerations of the image sampling procedure. Haralick [10] on the other hand has proposed an edge-detection scheme which draws on a facet model. These methods are effectively non-linear filtering techniques which produce an alternative characterisation of the edge information.

Spacek [2] has extended the surface modelling idea to the gradient magnitudes. The method is effectively a hybrid non-maximum suppression technique. Sets of twelve neighbouring gradient magnitudes are represented by a second order surface the parameters of which are determined by a least-squares fit. The coefficients of the polynomial surface are used to determine the maximum value of the derivative of the image luminance function in the direction of the gradient of the pixel under consideration. By virtue of the fitting procedure, Spacek's method gives sub-pixel acuity in the detection of the gradient maxima.

## Labelling Techniques

The simplest postprocessing of the localised gradient maxima is performed using binary thresholding. The edge labels obtained by binary thresholding can be used as input to Hilditch skeletonisation [15] to produce edge structures of a single pixel width. In this way the labelled edge map can draw on connectivity information. Although Hilditch edge-thinning draws on a dictionary, it does not incorporate any evidential index for the existence of edge structures.

Canny [1] also draws on connectivity information in his multi-threshold hysteresis linking scheme. Edge-pixels are initially labelled if their response exceeds a high threshold value. Pixels laying above a weaker response threshold are then admitted provided they belong to edge segments which are connected to the initially labelled pixels. Hysteresis linking admits more evidence than Hilditch skeletonisation since it uses two context-sensitive thresholds.

## EXPERIMENTAL COMPARISON

The space available here is too limited to present an exhaustive comparison of the postprocessing techniques listed above. We shall concentrate on demonstrating the robustness of the various approaches to image noise. We will not attempt to demonstrate the relative performance of the techniques on different filtered characterisations of the the image data. For the purposes of

experimentation we have therefore used noisy synthetic and natural images.

Figures 1 and 2 show the labellings obtained by the discrete and probabilistic relaxation methods for a synthetic image in which the signal to noise ratio is 25:8; both labellings are relatively free from contamination by extraneous noise artefacts and display good connectivity properties. For comparison, Figures 3 and 4 show the gradient maxima obtained using non-maximum suppression and Spacek's surface model. The output of non-maximum suppression is very sensitive to image noise and contains many isolated edge points; there is little to be gained from further processing. Spacek's result certainly provides a basis on which to improve using labelling techniques; since the edge response is already localised, there is no point using skeletonisation. Figures 5 and 6 show the results obtained by applying binary thresholding and hysteresis linking. The important point to note about these labellings is that although hysteresis linking removes isolated noise pixels, it has a tendency to encourage extraneous edge structure.

Figures 7, 8, 9, and 10 show the results obtained when probabilistic relaxation, discrete relaxation, Canny's algorithm and Spacek's surface model are applied to a noisy natural image. The same observations hold as in the synthetic case. The probabilistic relaxation method appears most robust and extracts the most meaningful connected edge structures.

## CONCLUSIONS

Dictionary-based relaxation methods provide a powerful way of objectively quantifying the contextual information pertaining to the edge postprocessing problem. Whereas alternative concepts draw on limited contextual information restricted either to the measurement or to the structural domain, the relaxation approaches simultaneously exploit both kinds of information. Experimental studies show that the probabilistic relaxation method is more robust to raw noise than the alternatives. This has the benefit that it can be applied directly to raw edge information without the need to filter the effects of noise with the consequent band-limitation of image features. In noise-free images most of the post-processing techniques perform in a satisfactory way.

## REFERENCES

- 1 Canny, J.F. "A Computational Approach to Edge Detection," *IEEE PAMI*, **8**, 679-700, 1986.
- 2 Spacek, L.A. "Edge detection and motion detection," *Image and Vision Computing*, **4**, 43-56, 1986.
- 3 Kittler J. and E. R. Hancock, "Combining Evidence in Probabilistic Relaxation," *International Journal of Pattern Recognition and Artificial Intelligence*, **3**, 29-52, 1989.
- 4 Hancock E.R. and J.Kittler, "Edge Labelling using Dictionary-based Relaxation," *Submitted for publication*, 1988.
- 5 Hancock E.R. and J.Kittler, "Discrete Relaxation," *Submitted for publication*, 1989.
- 6 Waltz, D.L. "Understanding line drawings of scenes with shadows" in *The Psychology of Computer Vision*, ed. P.H. Winston, McGraw-Hill, New York, 1975.
- 7 Rosenfeld, A., R.A. Hummel and S.W. Zucker, "Scene labeling by relaxation operations," *IEEE SMC*, **6**, 420-433, 1976.
- 8 Hummel, R.A. and S.W. Zucker, "On the foundations of relaxation labeling processes," *IEEE PAMI*, **5**, 267-287, 1983.
- 9 Nalwa, V. "On Detecting Edges," *IEEE PAMI*, **8**, 701-714, 1986.
- 10 Haralick R.M. and L.T.Watson, "A facet model for image data," *Computer Vision, Graphics and Image Processing*, **15**, 113-129, 1981.
- 11 Faugeras O.D. and M. Berthod, "Improving consistency and reducing ambiguity in stochastic labeling: an optimization approach," *IEEE PAMI*, **3**, 412-424, 1981.
- 12 Zucker, S.W., R.A. Hummel and A. Rosenfeld, "An application of relaxation labelling to line and curve enhancement," *IEEE Trans. Computers*, **C-26**, 394-403, 1977.
- 13 Geman S and D Geman, "Stochastic Relaxation, Gibbs Distributions and Bayesian Restoration of Images," *IEEE PAMI*, **6**, 721-741, 1984.
- 14 Haralick R.M. and J.S.J Lee, "Context dependent edge detection" *Proc. 9th ICPR, Rome, Vol 1*, 203-207, 1989.
- 15 Hilditch, J. "Linear skeletons form square cupboards," *Machine Intelligence*, **6**, 403-420, 1969.
- 16 Rosenfeld, A. and M. Thurston, "Edge and curve detection for visual scene analysis," *IEEE Trans. Computers*, **C-20**, 562-569, 1971.

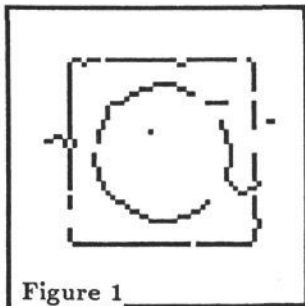


Figure 1

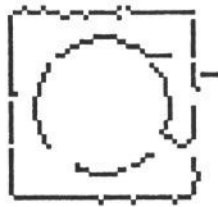


Figure 2



Figure 3



Figure 4

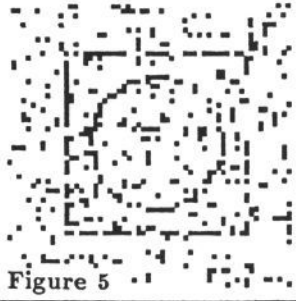


Figure 5



Figure 6

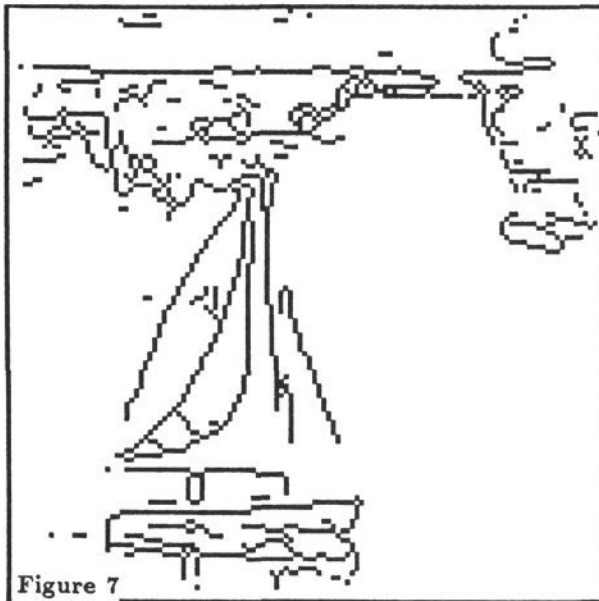


Figure 7

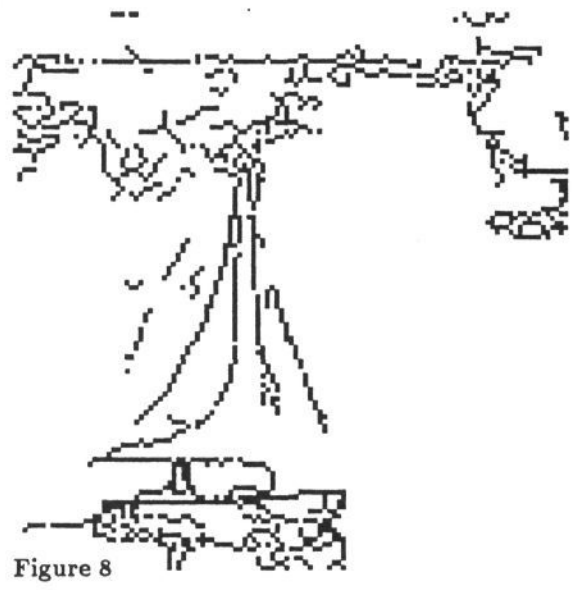


Figure 8

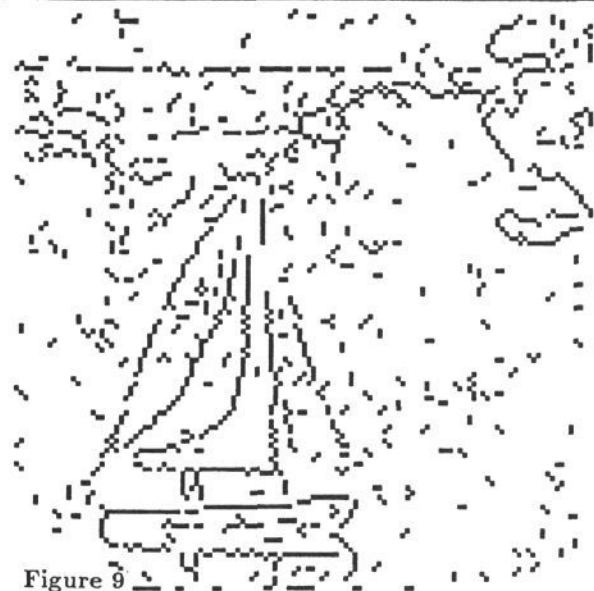


Figure 9



Figure 10