

Edge-Labeling Using Dictionary-Based Relaxation

EDWIN R. HANCOCK AND JOSEF KITTLER, MEMBER, IEEE

Abstract—We present an improved application of probabilistic relaxation to edge-labeling. The improvement derives from the use of a representation of the edge-process that is internally consistent and which utilizes a more complex description of edge-structure. The particular novelty of the application lies in the use of a dictionary to represent permitted labelings of the entire context-conveying neighborhood of each pixel. This approach is to be contrasted with the use of approximate factorizations which have been employed in previous applications to decompose the neighborhood into object-pairs. We give details of the dictionary approach and the related representation of the edge-process. A comparison with other edge-postprocessing strategies is provided. This leads us to conclude that the dictionary-based approach is a powerful edge-postprocessing tool. It relaxes the demands on the level of filtering that has to be applied to cope with image noise with the benefit of reduced blurring of fine image features.

Index Terms—Contextual classification, edge-labeling, probabilistic relaxation.

I. INTRODUCTION

PROBABILISTIC relaxation is a well known technique for labeling image entities. It relies on iteratively updating the distribution of available probability over a label set for each of the image entities; the crucial computational ingredient is a support function which combines evidence from the context-conveying neighborhood and incorporates prior knowledge of the structure of the labeling task in-hand. One of the earliest applications of probabilistic relaxation to labeling problems of realistic complexity was concerned with edge-enhancement. The application was not widely adopted as a practical edge-detection strategy for reasons of limited performance such as nonuniform convergence and the inability to enhance a realistic variety of edges. It was subsequently shown that these limitations resulted from internal inconsistencies in the specification of the relaxation scheme [13], the heuristic nature of the updating process [5], and a limited capacity to accurately represent edge-processes. Recent work has aimed at resolving the problems of consistency and improving representational capacity [3], [12], [14].

The rejection of probabilistic relaxation as a means of edge-labeling may therefore be regarded as premature. The objective of this paper is to demonstrate that the early limitations of probabilistic relaxation have been overcome

and that it can be used to label edges with a comparable, if not improved, performance to the most advanced existing edge-detection algorithms. In particular we will show the advantages to be gained from using a richer representation of the edge-process.

A. Edge-Labeling

Edge-labeling is acknowledged as an important early stage in computer vision. The correspondence between edge tokens in different images is used as the basis of several approaches to stereoscopic reconstruction of depth and also to deriving information about rigid body motion; edge-based descriptions are also used for model matching. In each of these applications, the robustness of the edge-finding process to noise is of vital importance. Robust edge-detection has therefore been a goal of computer vision for almost three decades. The task is a complex one. A complete physical description of the detection process would involve models of the image sensing device, scene illumination effects and of the object-geometry of edge-features. For example, boundary discontinuities between objects in uniformly illuminated scenes are characterized by step-profiles in the image luminance function. When sampled by a realistic sensing device, the step-profile will be blurred due to finite resolving power of the optical components and be subject to thermal noise contamination from the electronic components. Further complications are introduced when illumination gradients are present and when the location of nonstep edge types is attempted.

Most authors have confined their attention to the detection of restricted types of edge-structure. For instance, Canny [16], [17] and Spacek [9] have concentrated on the optimal characterization of step-profiles in the presence of noise using first derivatives of the image luminance function. The important conceptual stages in the edge-labeling task are as follows.

- The image is first filtered to smooth the effects of noise and, in some applications, produce a multiscale representation of image data. The filtered image luminance function is then differentiated with a view to finding significant edge-features. This procedure amplifies any residual noise-component present in the luminance function. Both Canny and Spacek derive optimal filters for the detection of step edges in the presence of noise. Spacek's filter satisfies the optimality criteria of maximum signal-to-noise ratio, maximum localization of the gradient-maxima associated with idealized edges and minimum peak density. The computation of directional derivatives is

Manuscript received May 17, 1988; revised August 16, 1989. Recommended for acceptance by C. Brown.

E. R. Hancock is with SERC Rutherford Laboratory, Chilton, Didcot, Oxfordshire, OX11 0QX, England, and the Department of Electronic and Electrical Engineering, University of Surrey, Guildford, GU2 5XH, England.

J. Kittler is with the Department of Electronic and Electrical Engineering, University of Surrey, Guildford, GU2 5XH, England.

IEEE Log Number 8931854.

achieved by first applying a circularly symmetric realization of the optimal filter and subsequently differentiating in the required orientation using operators of the smallest possible mask size. Canny, who originally formulated the above optimality criteria, resorts for computational reasons to the use of a suboptimal Gaussian filter. Although the directional derivatives are computed in a single stage using directional masks, the edge-filtering process is operationally identical to Spacek's method. In addition to smoothing the effects of noise, both filters have a tendency to remove genuine high-frequency edge-features such as corners.

- Having obtained the gradient vectors of the image luminance function, it is necessary to locate local maxima of the gradient magnitude. Both authors use nonmaximum suppression techniques for this purpose. Canny performs a simple interpolation of the gradient information in a 3×3 window to determine the direction and magnitude of the local gradient-maxima. Spacek performs a least-squares second-order polynomial fit to twelve neighboring gradient magnitudes. The coefficients of the polynomial are used to determine the maximum value of the derivative of the image luminance function in the direction of the gradient. By virtue of the fitting procedure, Spacek's method gives subpixel acuity in the detection of the gradient-maxima.

- Edge-pixels are labeled on the basis of the information provided by the localized gradient-maxima of the luminance function. In Spacek's detector this can be achieved using a simple binary thresholding procedure. Canny uses multithreshold hysteresis linking which draws on information concerning edge-connectivity. Edge-pixels are initially labeled if their response exceeds a high-threshold value. Pixels lying above a weaker response threshold are then admitted provided they belong to edge-segments which are connected to the initially labeled pixels. Finally, unconnected high-response pixels are deleted.

Canny and Spacek use first derivatives of the filtered image luminance function as the basis of their edge-detection methods. Other authors characterize the raw edge-information in different ways. On the basis of psychophysical observations, Marr and Hildreth [19] argue for the use of zero crossings of the Laplacian of Gaussian operator. The properties of zero crossings ensure that the detected edges are connected. Since the differential operator is second-order its noise amplification effects are severe unless it draws on a large support mask; this limits its capacity to faithfully detect high-curvature edge-features. Rather than using differential operators Nalwa [18] directly interprets the information in the image luminance function by fitting profiles in restricted windows. The form of these profiles is motivated by considerations of the image sampling procedure. Haralick [20] on the other hand has proposed an edge-detection scheme which draws on a facet model of the image luminance function.

The approaches listed above are concerned with the way that raw edge-information in the image luminance function is characterized. In this paper we are interested in the

later stages of edge-detection, i.e., the postprocessing of the raw or filtered edge-information. The task is concerned with locating and labeling consistent edge-structures given confused or noisy response information. In practice it invariably draws on various sources of contextual information such as the gradient vectors available in a neighborhood of the image or the connectivity of candidate edge-structures. For instance Canny performs non-maxima suppression by interpolating gradient vectors from an image neighborhood and draws on connectivity information in the hysteresis linking stage. Spacek, on the other hand, imposes analytic continuity on the gradient magnitudes by fitting a second-order polynomial surface to achieve nonmaximum suppression. Although these schemes are certainly effective solutions to the problem, they do not necessarily make best use of the available contextual information. It is in this respect that the contextual labeling schemes can be profitably applied to the edge-detection problem.

B. Relaxation Processes

A host of contextual labeling strategies are described in the literature. The bulk of these is encompassed by the term relaxation. The common aim is to improve the consistency of object classification. This is achieved by utilizing observational information in the form of measurements pertaining to objects and prior knowledge of the constraints which apply in a particular labeling 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 neighboring objects. Both types of relaxation process have received considerable attention in the literature and have a variety of realizations.

The original discrete relaxation algorithm was developed by Waltz [21]. This work is significant for two reasons. First, it showed how globally consistent label configurations could be obtained using a constraint filtering algorithm. The second aspect is more pertinent to the work reported here; it showed how knowledge relating to a particular labeling application could be represented by a dictionary of symbolic constraints applying to groups of objects. Unfortunately, the constraint filtering algorithm had a number of shortcomings. It proved difficult to regulate for general labeling problems and by virtue of the purely symbolic representation of the labeling task, failed to admit the wealth of available observational information.

It was the need to overcome these shortcomings that led Rosenfeld, Hummel, and Zucker [1] to the probabilistic relaxation method. However, a host of difficulties have subsequently been identified and are pursued in the literature. In particular Hummel and Zucker [5] have addressed the problems associated with the heuristic nature of the update procedure. Recent works by Kittler and Foglein [13] and by Kittler and Hancock [6] have considered ways of removing internal inconsistencies of speci-

fication and improving representational capacity. Kittler and Illingworth [2] provide a recent review of the extensions and applications of the probabilistic relaxation method.

The discrete relaxation method has itself been extended in a number of ways. Recent work has aimed at incorporating observational information by regarding the labeling task as maximum *a posteriori* probability estimation. Algorithms falling into this category include the stochastic relaxation scheme of Geman and Geman [22] and Besag's iterated conditional modes [26]. Both algorithms have been applied to low-level image interpretation problems and draw on Markov models of the label process. In fact, the use of Markov models in connection with image segmentation has proved a particularly fruitful area of study. The problem has been addressed by Derin, Elliott, Cristi, and Geman [23], and, by Hanson and Elliott [24]. The Markov model used by Geman and Geman [22] simultaneously accounts for region and edge-processes in image segmentation. None of this work is concerned purely with the postprocessing of edge-information. For instance, even though Geman and Geman [22] incorporate a boundary process into their model, it is difficult to assess its effectiveness independently of the region process.

An interesting application of Markov modeling to edge-postprocessing has been reported by Haralick [25]. The method draws on a facet model characterization of edge-response with the label process drawing on a Markov chain concept. The Markov chain allows label probabilities to be computed recursively by distinguishing between the causal past and future of continuous edge-structures. By virtue of the chain indexing of continuous edges the label process can be expressed using pairwise relationships between adjacent edge-labels; a separate decision is required to assign the edge-direction.

Although all of these extensions of the relaxation concept admit observational information, they have not drawn on Waltz's dictionary idea [21] to model the label process. We have addressed this issue in [6] by detailing a probabilistic relaxation method which draws directly on the dictionary concept to combine evidence for label assignments. The dictionary method has a vastly improved capacity to represent highly structured labeling applications. Our aim in this paper is to apply the dictionary-based probabilistic relaxation method to the postprocessing of edge-information and to demonstrate its performance advantages. Regardless of any merits, the Markov methods described above are not directly comparable to the dictionary-based approach.

C. Edge-Labeling by Probabilistic Relaxation

Previous applications of probabilistic relaxation labeling to edge-detection and curve enhancement have invariably exploited prior knowledge of the permitted label assignments to pairs of pixels [4], [8], [10]. The use of such limited prior knowledge effectively corresponds to factorization of the entire context-conveying neighborhood [3], [12], [14] with the inevitable consequence of loss of edge-modeling capacity. In this paper we will make

use of the permissible labelings for the entire context-conveying neighborhood. These labelings are represented by a relaxation. When used in conjunction with probabilistic relaxation, the dictionary approach has the following advantages.

- It avoids the necessity to approximate the world model by marginal probabilities for subunits (e.g., pixel pairs) with a corresponding loss of representation and in some case admitting nonphysical labelings. The method therefore allows a much richer representation of the structure of the world model. Factorization schemes are associated with limiting assumptions which restrict the capacity of the world model to adequately cope with realistic labeling problems.

- It represents a considerable improvement in computational efficiency. The complexity of the support function associated with factorization schemes is an exponential function of the size of the label set and the number of objects in each representational unit. The complexity of the dictionary-based support function, on the other hand, is determined by the number of permissible labelings. For most highly structured labeling applications, the length of the dictionary is very much smaller than the number of combinatorially occurring configurations.

It must therefore be stressed that the novelty of the application described in this paper lies in the use of the dictionary concept to represent edge-structure. The edge-labeling problem is highly amenable to this approach since the permitted labelings of context-conveying neighborhoods are determined by connectivity on the pixel lattice and are therefore easily compiled in a dictionary. This property allows a more sophisticated model of the edge-process to be used. It should also be noted that the support functions used in the previous applications have displayed internal inconsistency; for instance, after an initial subjective improvement, the image labeling deteriorates displaying effects such as edge-thickening. The support functions used in our formulation are internally consistent and we have demonstrated that the relaxation process converges uniformly upon a labeling which is objectively consistent and is in subjective agreement with the image contents [3].

The use of dictionary-based probabilistic relaxation contrasts with the postprocessing approaches used by Canny and Spacek in the following ways.

- Initial probabilities for edge- and non-edge-pixel classes are used to represent gradient information. The gradients can either be derived from the raw image luminance function or from the output of a filter. Following Spacek, the gradient information is computed using differencing operators based on the smallest possible support masks; this choice clearly separates the effects of filtering and differentiation. The label probabilities model the effects of noise in the non-edge-class; this is a novel feature of the method since it does not require a detailed description of the appearance of edge-profiles in the image luminance function. The method is not limited in its capacity to cope with high levels of noise by the small size of the support masks used for differentiation; it can equally

well be applied to the output of noise suppressing filters of large support mask. It should be noted that the optimal filters used by Canny and Spacek are solely concerned with the detection of step-profiles in the presence of noise.

- Prior knowledge of the labelings permitted by edge-connectivity is represented in a dictionary. The salient properties of the label process for edge-labeling are encapsulated by a dictionary for the 3×3 neighborhood. There is little representational advantage to be gained from the use of dictionaries for larger neighborhoods. In the work of Spacek there is no attempt to draw on this type of symbolic constraint to improve the connectivity of detected edges. Canny draws on an implicit dictionary concept by measuring an index of connectivity during hysteresis linking.

- The evidence combining procedure accumulates support for the edge- and non-edge-labels from the items in the dictionary. The procedure is applied recursively until the label probabilities converge on a hard labeling which has a unique interpretation. In doing this support is incorporated from an increasingly large neighborhood; the final label probabilities eventually reflect support from a neighborhood which is of much greater size than that used by the evidence combining formula. In performing a polynomial fit to gradient magnitudes Spacek effectively draws on evidence from an image neighborhood. Canny on the other hand works with a symbolic representation of the labeling task which is derived from the gradient magnitudes for single pixels using a set of thresholds.

According to the approach described above, raw gradient information is regarded as evidence that should be combined with our expectations based on prior knowledge. By exploiting the rich context of prior knowledge to its full extent the need for noise suppressing filtering is minimized. This reduces the band limitation of features in the image luminance function. Although we have concentrated on the postprocessing of information derived from first-difference operators, the approach could also be applied to alternative characterizations of the raw edge-information (e.g., Nalwa [18], Haralick [20]) provided that probabilistic models of the measurement process are available.

The outline of this paper is as follows. In Section II, we review the formulation of probabilistic relaxation necessary for the edge-labeling application. Section III contains details of the computational realization of the relaxation approach. In Section IV, we present an evaluation of the performance of dictionary-based probabilistic relaxation for edge-labeling and provide a comparison to other approaches.

II. DICTIONARY-BASED PROBABILISTIC RELAXATION

An important task in scene interpretation is the global labeling of objects in an unambiguous and consistent way. The algorithms which prove most successful in this respect invariably draw on information concerning object-context and utilize prior knowledge of the structure of the labeling problem in-hand. Probabilistic relaxation is a technique that embodies these ingredients. First formu-

lated in a seminal paper by Rosenfeld, Hummel, and Zucker [1], the technique has received sustained interest over the past decade.

The problem addressed in probabilistic relaxation is that of finding the class identity θ_j of each object in a network $a_j, j = 1, N$ according to a set of class labels Ω . According to the philosophy of probabilistic relaxation, each object simultaneously admits every label in the set with probabilities that reflect the evidence. Central to the approach suggested by Rosenfeld, Hummel, and Zucker is the iterative updating of the label probabilities for each object using an evidence combining formula. In order to facilitate the development of the relaxation formalism we shall use the notation ω_{θ_j} to denote a particular realization of the class identity of object a_j . At the n th iteration of the relaxation process, the label probability for this assignment is $P^{(n)}(\theta_j = \omega_{\theta_j})$. The updating of the label probability draws on contextual information through a support function for the object-label assignment, i.e., $Q^{(n)}(\theta_j = \omega_{\theta_j})$. The original evidence combining formula proposed by Rosenfeld, Hummel, and Zucker, is

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

The explicit specification of the support function was a prerequisite of the original RHZ scheme. As such it tended to be based on inconsistent assumptions concerning the way in which contextual support is quantified. In [3], [12], and [14] we have addressed the problem of combining evidence in probabilistic relaxation in a framework which does not commence with a specification of the support function. Instead, we commence by specifying a set of probabilistic relationships which have distinct and precise meaning. Since the specification dictates the representation of world-knowledge, the relaxation mechanism is internally consistent. The probabilistic representation has the following constituents

- The initial label probabilities

$P^{(0)}(\theta_j = \omega) = P(\theta_j = \omega | x_j)$ are derived from the raw measurements for objects $x_j, j = 1, N$. By recursively combining evidence we improve the estimate of the label probabilities by incorporating contextual information. We regard the updating of label probabilities as an implicit filtering of the raw observations. Since at each update the current probabilities reflect contextual evidence from increasingly large neighborhoods, the repeated updating would eventually achieve the desired objective of drawing global contextual information from the complete network. This is the essence and philosophy behind probabilistic relaxation algorithms.

- The joint prior $P(\theta_j = \omega_{\theta_j}, \theta_l = \omega_{\theta_l}, \forall l \in I_j^*)$ is a probabilistic representation of the world-model of label structure for the objects in the context-conveying neighborhood of object j , i.e., the objects with index-set I_j^* . The single object priors $P(\theta_l = \omega_{\theta_l})$ are obtained from the joint priors by applying the axiomatic properties of joint probability.

With this framework, the support function becomes [3], [12], [14]

$$Q^{(n)}(\theta_j = \omega_{\theta_j}) = \frac{1}{P(\theta_j = \omega_{\theta_j})} \sum_{\omega_{\theta_l} \in \Omega} \left\{ \prod_{l \in I_j^*} \frac{P^{(n)}(\theta_l = \omega_{\theta_l})}{P(\theta_l = \omega_{\theta_l})} \right\} \cdot P(\theta_j = \omega_{\theta_j}, \theta_l = \omega_{\theta_l}, \forall l \in I_j^*). \quad (2)$$

The summation over object-label assignments to the neighborhood of directly interacting objects, i.e., $\sum_{\omega_{\theta_l} \in \Omega}$, implies that the evidence combining formula is of exponential complexity. However, in the following section we shall demonstrate that in practice the actual complexity, far from being exponential, may compare favorably to factorization approaches which implement support calculation in polynomial time [1], [4], [6], [8], [10].

B. Support Derived from Dictionary Look-Up

In order to render the evidence combining formula given in (2) useful for realistic labeling problems, it is necessary to derive support functions of at most polynomial complexity. One approach to reducing the computational complexity of the support function is to factorize the joint prior into marginal probabilities for smaller subunits of the context-conveying neighborhood. The need to maintain the internal consistency of the relaxation scheme dictates the size of subunits. For instance the four pixel neighborhood must be factorized in terms of pairs of objects while the eight pixel neighborhood is factorized into groups of four pixels; the complexity of the resulting support functions are of order m^2 and m^4 , respectively [5], [14].

Alternatively, and central to the approach adopted in this paper, the exponential rise in computational expense with the size of representational units can be offset by the following observation: although the joint prior is potentially of exponential complexity, in many realistic applications the number of permissible configurations of labels in the contextual neighborhood is relatively small. This restriction in complexity results from the fact that many labeling problems are in fact highly structured; this will be demonstrated for the case of edge-labeling. As a consequence it becomes feasible to make an exhaustive compilation of the permissible configurations and to list them in a dictionary. This idea was central to the discrete relaxation method of Waltz [21].

As we shall demonstrate in Section IV, for the edge-labeling application the compilation of the dictionary is very straightforward; it consists of all labelings of a 3×3 neighborhood permitted by eight connectivity on a pixel lattice. The highly structured nature of the application becomes clear when it is noted that the dictionary contains only 181 of the 5^9 combinatorial possibilities. The implementation of the dictionary-based method is very simple for another reason; every object neighborhood has an identical dictionary. Waltz demonstrated that the dictionary was capable of encapsulating the label process for more complex applications. For interpreting line drawings of blocks-world scenes the dictionary is a compila-

tion of the constraints applying at line junctions. These junctions are of varying types and topology. The dictionary therefore contains items of different length and has sections for different junction topologies.

Although the power of the dictionary representation of constraints was evident from the work of Waltz, it has not been adopted in work on probabilistic relaxation. In a recent paper we have addressed this issue by detailing an evidence combining approach to probabilistic relaxation which utilizes the dictionary concept [3]. The method represents an improvement on existing techniques since it draws on the labeling of the entire context-conveying neighborhood of each object. Our aim here is to review the formalism necessary to apply the method to the edge-labeling application.

If the number of permissible labelings Z_j for each of the object neighborhoods I_j is small, then they can be realistically listed in a dictionary \mathcal{D}_j . The dictionary for each object neighborhood is subdivided into sections $\mathcal{D}_j(\omega_{\theta_j})$ which have length $Z_j(\omega_{\theta_j})$ and which are indexed according to the value center-object label ω_{θ_j} . Let $\omega_{\theta_l}^k$ denote the label on the object a_l , $l \neq j$, required by the k th entry in the section $\mathcal{D}_j(\omega_{\theta_j})$ of the dictionary. Accordingly, the k th section entry in the dictionary is a set of labels defined over the context-conveying neighborhood of object a_l , which we denote by

$$\Lambda_j^k(\omega_{\theta_j}) = \{ \theta_l = \omega_{\theta_l}, \theta_l = \omega_{\theta_l}^k, \forall l \in I_j^* \}. \quad (3)$$

Having defined the dictionary concept it remains to show how the associated joint priors for the permissible labelings may be modeled. In the work reported here we apportion the total available probability mass among the physical dictionary items. For the edge-labeling application we adopt a uniform distribution of probability mass. This need not necessarily be the case; in some applications it may be desirable to encourage some configurations at the expense of others. Accordingly, the probability measure associated with any physically impossible configuration $\{ \theta_l = \omega_{\theta_l}, \forall l \in I_j \} \notin \mathcal{D}_j(\omega_{\theta_j})$ is zero, i.e.,

$$P(\{ \theta_j = \omega_{\theta_j}, \theta_l = \omega_{\theta_l}, \forall l \in I_j^* \} \notin \mathcal{D}_j(\omega_{\theta_j})) = 0. \quad (4)$$

This dictionary model of the label process results in a critical simplification of the support function in (2). The implied multiple sum over label assignments to the neighborhood of directly interacting objects can be replaced by a sum over labelings in the subsection of the dictionary $\mathcal{D}_j(\omega_{\theta_j})$ and the support function becomes

$$Q^{(n)}(\theta_j = \omega_{\theta_j}) = \frac{1}{P(\theta_j = \omega_{\theta_j})} \cdot \sum_{k=1}^{Z_j(\omega_{\theta_j})} \left\{ \prod_{l \in I_j^*} \frac{P^{(n)}(\theta_l = \omega_{\theta_l}^k)}{P(\theta_l = \omega_{\theta_l}^k)} \right\} \cdot P(\Lambda_j^k(\omega_{\theta_j})). \quad (5)$$

This support function allows us to use knowledge concerning the label assignments to the full context-convey-

ing neighborhood; in the next section we will demonstrate how it can be used to express a more sophisticated model of edge-processes.

III. EDGE-LABELING USING DICTIONARY-BASED SUPPORT

The observation that the behavior of the physical edges is governed by connectivity on a pixel lattice suggests that their behavior can be successfully encapsulated in a dictionary. Provided it is possible to characterize edge-information using a set of initial label probabilities, then dictionary-based probabilistic relaxation may provide a useful framework for combining evidence to label edge-pixels. Evidence is recursively combined by drawing on the dictionary of permissible labelings in the context-conveying neighborhood until a hard labeling is obtained. The way in which the evidence combining approach draws on the available contextual information can be contrasted with Canny's hysteresis linking scheme or Spacek's polynomial fitting of gradient magnitudes. Neither approach simultaneously combines evidence and exploits knowledge concerning edge-structure. Although hysteresis linking does effectively draw on a dictionary to determine connectivity, it does not combine evidence. Edge-pixels are labeled on the basis of their connectivity and the value of their gradient magnitude. Spacek's procedure draws on a polynomial fit to the gradient magnitudes in a local region of the image. The gradient-maxima determined in this way have increased connectivity by virtue of the analytic nature of the fitting procedure. However, the method does not draw on any kind of dictionary concept; the final hard-labeling of edge-pixels is performed using a thresholding operation which makes no use of prior knowledge of edge-structure.

In order to apply the support function given in (5) to edge-postprocessing we must provide the following computational ingredients.

- A set of class labels to describe the edge-process.
- Initial label probabilities derived from the image luminance function.
- An appropriately structured dictionary of edge-labeling possibilities.
- *A priori* probabilities for the dictionary items and single label probabilities.

In Section III-A we describe the meaning of the labels used to represent the edge-process, Section III-B contains a discussion of the procedure used to calculate initial label probabilities, and Section III-C gives details of the dictionary and the inference of *a priori* probabilities.

A. Class Labels for the Edge-Process

The choice of class labels used to represent edges involves a compromise between exploiting the full structural richness of permissible neighborhood labelings and the computational complexity of the resulting evidence combining formula. For the implementation described here we have chosen to use a five label set. This consists of the four labels which represent edge-propagation par-

allel and antiparallel to the two pixel lattice axes and a non-edge-label. The notation used to denote this label set is $\Omega = \{\leftarrow, \uparrow, \rightarrow, \downarrow, \phi\}$ where the arrows are perpendicular to the intensity gradients associated with the edges and ϕ is the non-edge-label. As an example, \rightarrow denotes a pixel belonging to an edge which is associated with a boundary that is parallel to the horizontal axis of the pixel lattice.

B. Initial *A Posteriori* Probabilities

The role of the initial *a posteriori* probabilities is to model the affinity of raw measurements derived from a gray-scale image to the class labels used to represent the edge-process. This section describes the model adopted and its calculational realization in terms of the derivatives of the image luminance function.

If the initial gray-level image is composed of regions of uniform intensity bounded by step edges, then the corresponding directional intensity gradients will be characterized by delta functions of differing amplitude. In reality, the regions in the gray-level image will neither be of uniform intensity nor bounded by step edges. In particular the individual pixel gray-levels may be subject to random noise and the boundaries of regions may extend over several pixels. As a result, the intensity gradient will not be simply characterized by delta functions, but will consist of many complex features including isolated noise pixels, ramps, and ridges; any model that attempted to describe such a variety of edges would contain many parameters. In order to avoid bias in favor of any particular type of edge-profile, we have concentrated on modeling the noise processes. The model assumptions concerning the edge-process are independent of the gray-scale profile and are expressed only in terms of the admissible edge-configurations which are described in Section III-C.

A conventional model of the noise generation process has been adopted here, that is, that the noise occurring in the pixel gray-levels is assumed to be additive and is generated by a Gaussian probability distribution function with zero mean and standard deviation σ . The intensity gradients in the pixel lattice directions are derived from the mask shown in Fig. 1 using the following first differences:

$$\begin{aligned} c_1 &= g_1 - g_0 \\ c_2 &= g_2 - g_0. \end{aligned} \quad (6)$$

In vector form the above transformation of gray-scale values can be expressed as $c = A^T g$ where $c = (c_1, c_2)^T$, $g = (g_0, g_1, g_2)^T$ and

$$A = \begin{pmatrix} -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (7)$$

If the pixel g_0 belongs to the non-edge-class, then each of the pixels g_0, g_1 , and g_2 must belong to the same image region; this implies that the pixel gray-levels are subject

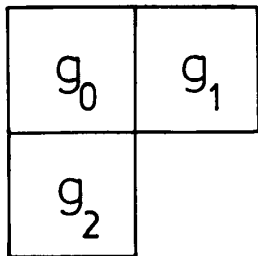


Fig. 1. Gray-level values used to calculate partial derivatives.

to identical distribution functions which have the same mean and standard deviation. Consequently, c_1 and c_2 will be of zero mean.

Under the assumption that the noise component in the pixel gray-level value acts as an independent random

$$p(c_1, c_2 | \theta_0 = \phi) = \frac{1}{2\pi \sqrt{4(s-u)(s-t) - (t+u-s-v)^2} \sigma^2} \cdot \exp \left\{ -\frac{(s-u)c_1^2 + (s-t)c_2^2 + (t+u-s-v)c_1c_2}{\{4(s-u)(s-t) - (t+u-s-v)^2\} \sigma^2} \right\}. \quad (16)$$

variable, the joint conditional density of \mathbf{g} is normal with variance-covariance matrix

$$\Sigma = \begin{pmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix}. \quad (8)$$

However, as we indicated in the Introduction, the characterization of edge-information frequently involves a filtering stage. In this case the gray-scale values will be correlated and in consequence the variance-covariance matrix will no longer be diagonal. Suppose that the raw image has been processed using the $N \times N$ filter

$$\mathbf{F} = \begin{pmatrix} a_{1,1} & \cdots & a_{1,N} \\ \cdots & \cdots & \cdots \\ a_{N,1} & \cdots & a_{N,N} \end{pmatrix}. \quad (9)$$

The variance-covariance matrix for the filtered gray-scale values is as follows

$$\Sigma = \begin{pmatrix} s\sigma^2 & t\sigma^2 & u\sigma^2 \\ t\sigma^2 & s\sigma^2 & v\sigma^2 \\ u\sigma^2 & v\sigma^2 & s\sigma^2 \end{pmatrix} \quad (10)$$

where the covariance elements resulting from the correlations introduced by the filtering operation are

$$s = \sum_{i=1}^N \sum_{j=1}^N a_{i,j}^2 \quad (11)$$

$$t = \sum_{i=1}^N \sum_{j=2}^N a_{i,j} \cdot a_{i,j-1} \quad (12)$$

$$u = \sum_{i=2}^N \sum_{j=1}^N a_{i,j} \cdot a_{i-1,j} \quad (13)$$

$$v = \sum_{i=2}^N \sum_{j=2}^N a_{i,j} \cdot a_{i-1,j-1}. \quad (14)$$

In both the uncorrelated and correlated cases, the vector of first differences \mathbf{c} is obtained by the linear transformation of the gray-scale values $\mathbf{c} = \mathbf{A}^T \mathbf{g}$. As a consequence \mathbf{c} is normal with distribution function

$$p(\mathbf{c} | \theta_0 = \phi) = \frac{1}{2\pi \sqrt{|\tilde{\Sigma}|}} \exp \left\{ -\frac{1}{2} \mathbf{c}^T \tilde{\Sigma}^{-1} \mathbf{c} \right\} \quad (15)$$

where the transformed variance-covariance matrix $\tilde{\Sigma} = \mathbf{A}^T \Sigma \mathbf{A}$. After substitution for the transformation matrix the general expression for the class conditional density for the first differences is

In the case of raw unfiltered observations when $s = 1$ and $t = u = v = 0$, then the density function reduces to

$$p(c_1, c_2 | \theta_0 = \phi) = \frac{1}{2\pi \sqrt{3}\sigma^2} \exp \left\{ -\frac{c_1^2 + c_2^2 - c_1c_2}{3\sigma^2} \right\}. \quad (17)$$

In Section IV we report how the dictionary-based relaxation method can be used in conjunction with the output of a noise suppressing filter. For the purpose of experimentation, we use Spacek's [9] optimal filter. As a concrete example, when this filter is computed on a 5×5 support mask, the numerical values of the correlations s , t , u , and v result in the following density function

$$p(c_1, c_2 | \theta_0 = \phi) = \frac{1}{2\pi 1.4\sigma^2} \cdot \exp \left\{ -\frac{0.36(c_1^2 + c_2^2) - 0.067c_1c_2}{\sigma^2} \right\}. \quad (18)$$

It is interesting to note that as the mask size of the Spacek filter increases, the cross product of the directional derivatives has a diminishing effect and in the limit of large mask size the density has circular rather than elliptical symmetry.

Having derived the general conditional distribution function for the first differences and presented some specific instances, it remains to be shown how the initial label probabilities can be computed. The *a posteriori* probability of the non-edge-label is obtained from the conditional distribution function by applying the Bayes

rule

$$P(\theta_0 = \phi | c_1, c_2) = \frac{p(c_1, c_2 | \theta_0 = \phi)}{p(c_1, c_2)} P(\theta_0 = \phi). \quad (19)$$

Since the form of the mixture density function is not known, it is only possible to place a bound on the *a posteriori* probabilities; this is done by making the assumption that the density function corresponding to zero intensity gradient is a maximum, i.e., $p(c_1 = 0, c_2 = 0) \geq p(c_1 \neq 0, c_2 \neq 0)$. This assumption is quite realistic as the mixing proportions of edge-pixels will be very small in comparison to non-edge-pixels (i.e., the majority of pixels in an image will be non-edge-pixels). Moreover, in the vicinity of the origin, the mixture density will be dominated entirely by the conditional probability density which peaks at zero. The assumption, combined with the Bayes formula, leads to the following inequality:

$$\begin{aligned} & \frac{p(c_1 = 0, c_2 = 0 | \theta_0 = \phi)}{p(c_1 \neq 0, c_2 \neq 0 | \theta_0 = \phi)} \\ & \geq \frac{P(\theta_0 = \phi | c_1 = 0, c_2 = 0)}{P(\theta_0 = \phi | c_1 \neq 0, c_2 \neq 0)}. \end{aligned} \quad (20)$$

If the *a posteriori* probability of the non-edge-label is taken to be unity when the total edge-response is zero, then

$$P(\theta_0 = \phi | c_1 \neq 0, c_2 \neq 0) \geq \exp \left\{ - \frac{(s-u)c_1^2 + (s-t)c_2^2 + (t+u-s-v)c_1c_2}{\{4(s-u)(s-t) - (t+u-s-v)^2\} \sigma^2} \right\}. \quad (21)$$

In the work described here, we have taken the lower bound for the *a posteriori* probability of the non-edge-label; the residual probability which may be assigned to the four edge-labels is then

$$P_{\text{residual}} = 1 - \exp \left\{ - \frac{(s-u)c_1^2 + (s-t)c_2^2 + (t+u-s-v)c_1c_2}{\{4(s-u)(s-t) - (t+u-s-v)^2\} \sigma^2} \right\}. \quad (22)$$

Finally, the *a posteriori* probabilities for the four edge-labels are obtained by apportioning P_{residual} according to the relative signs and magnitudes of c_1 and c_2 :

$$\begin{aligned} P(\theta_0 = \downarrow | c_1, c_2) &= \frac{-c_1}{\Delta} P_{\text{residual}} && \text{if } c_1 \leq 0 \\ &= 0 && \text{otherwise} \\ P(\theta_0 = \uparrow | c_1, c_2) &= \frac{-c_2}{\Delta} P_{\text{residual}} && \text{if } c_2 \leq 0 \\ &= 0 && \text{otherwise} \\ P(\theta_0 = \rightarrow | c_1, c_2) &= \frac{c_1}{\Delta} P_{\text{residual}} && \text{if } c_1 \geq 0 \\ &= 0 && \text{otherwise} \end{aligned}$$

$$\begin{aligned} P(\theta_0 = \downarrow | c_1, c_2) &= \frac{c_2}{\Delta} P_{\text{residual}} && \text{if } c_1 \geq 0 \\ &= 0 && \text{otherwise} \end{aligned}$$

where

$$\Delta = |c_1| + |c_2|. \quad (23)$$

To summarize, the above procedure for estimating the initial label probabilities has the following properties.

- It uses the smallest mask size to estimate the derivatives of the image luminance function. This has the consequence of preserving features of high-spatial-frequency such as corners. The method used to calculate derivatives uses a simple 2*1 difference operator.

- It does not make any assumptions concerning the luminance profile of the edge-types that might be present in an image. The method does not assume that edges present step-profiles.

- It models the effects of noise through the joint conditional density function for the first differences. The method can process raw unfiltered edge-information.

The model described above should be contrasted with those adopted by Berthod and Faugeras [4] and by Zucker *et al.* [8] in their applications of probabilistic relaxation to edge-labeling. 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 mea-

surements originate from realistic edge or noise processes. We, on the other hand, have attempted to justify the way in which the initial probabilities are calculated.

However, we acknowledge that the problem is complex and that the model we adopt does have limitations. For instance, when the thermal noise contamination is low and the image noise is dominated by quantization errors, a better model would involve a uniform rather than Gaussian density. However, based on the experimental studies reported in Section IV-C, these limitations do not appear to impair performance.

C. The Dictionary and Associated Probabilities

The role of the dictionary is to encapsulate knowledge concerning the permissible labelings of the context-conveying neighborhood. Since we wish to exploit the greatest amount of evidence available for each pixel, we will use the eight pixel neighborhood shown in Fig. 2 to cal-

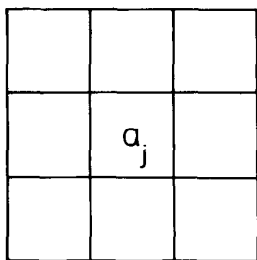


Fig. 2. Eight pixel context-conveying neighborhood.

culate support. We are therefore concerned with generating the dictionary of possible labelings of a 3*3 pixel window.

The dictionary must contain all physically reasonable labelings. It must not be biased toward particular types of edge-structure such as continuous straight lines or corners. To this end we have adopted a model of the edge-label process based on connectivity within the 3*3 pixel window. Briefly the properties of this model are as follows.

- Edges are one pixel wide.
- Permissible edges propagate continuously in one direction or undergo changes in the direction of $\pi/2$.

Examples of labelings that result from this model are shown in Fig. 3. The full dictionary has 181 items and the cardinalities of the sections are as follows:

$$\begin{aligned} Z(\leftarrow) &= Z(\uparrow) = Z(\rightarrow) = Z(\downarrow) = 17 \\ Z(\phi) &= 113. \end{aligned} \quad (24)$$

A further model assumption is necessary to infer the *a priori* probabilities for the dictionary items. In order not to bias the relaxation method in favor of particular edge-structures we have assumed that each dictionary item is equally likely, i.e.,

$$P(\Lambda^k(\omega)) = \frac{1}{Z} \quad (25)$$

where $\Lambda^k(\omega)$ is the k th item of the section $\mathcal{D}(\omega)$ of the dictionary. This method gives equal likelihood to all permissible edge-structures and does not, for instance, encourage straight edges at the expense of corners.

The *a priori* probabilities for single pixel label assignments were then inferred using the axiomatic property of joint probability, i.e.,

$$P(\theta_0 = \omega) = \sum_{k=1}^{Z(\omega)} P(\Lambda^k(\omega)). \quad (26)$$

This procedure for calculating the *a priori* probabilities differs from the approaches adopted by Berthod and Faugeras and by Zucker *et al.* in the following respects.

- It uses knowledge of labeling structure for the whole context-conveying neighborhood; the previous work only made use of *a priori* probabilities for pixel pairs. We therefore exploit more information concerning the structure of the edge-labeling application.

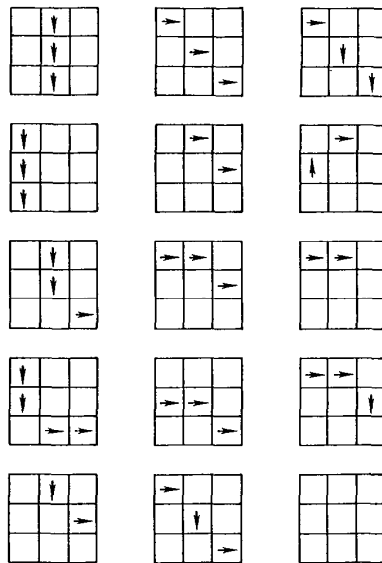


Fig. 3. Examples from the dictionary of neighborhood labelings.

- In the work of Zucker [8], the probabilities took the form of heuristic compatibility coefficients in the range -1.0 – 1.0 ; these coefficients were chosen to encourage continuous edges and suppress corners. A similar goal oriented approach was adopted by Berthod and Faugeras [4] who derived a set of probabilities with the objective of encouraging continuous straight edges. Our method does not exhibit these biases.

IV. APPLICATION OF THE EDGE-MODEL

This section describes experimentation undertaken to investigate the performance of the dictionary-based probabilistic relaxation algorithm for the edge-labeling process described in Section III. These experiments were intended to evaluate the robustness of the dictionary-based algorithm and to compare its performance with other approaches to edge-labeling.

A. Comparison to Other Methods

The following edge-labeling algorithms have been studied in order to assess the performance of the dictionary-based approach described in the previous section.

- Hilditch edge-thinning as an example of an incremental labeling algorithm applied to the thresholded edge-magnitudes [11]. The Hilditch algorithm also uses lists of neighborhood labelings and the comparison is made to show the advantages of using an evidence combining approach. This procedure does attempt to recognize label structure but it does not attempt to combine evidence.

- We have performed a detailed comparison with the edge-detector proposed by Spacek, since this represents one of the most theoretically consistent and practically successful approaches described in the literature. There are two aspects to the comparison to Spacek's work. First, we have compared the results of relaxation labeling to the

thresholding of fitted gradient magnitudes; the comparison uses identical differencing techniques in order to assess which method is most robust to noise. Second, we have compared the two labeling strategies for images convolved with filtering masks of different size in order to determine which method most successfully reconstructs edges for a given mask size.

- Although we have chosen Spacek's algorithm as the basic metric of comparison, we also present some experimental comparison to the work of Canny. In this respect we are concerned with comparing the nonmaximum suppression and hysteresis linking stages to dictionary-based probabilistic relaxation. The comparison is not extensive. We will suffice to show that Canny's postprocessing is only robust to noise if applied to a filtered representation of the image.

B. Imagery

Both synthetic and natural images have been used to compare the dictionary-based edge-labeling algorithm with the approaches described above. The synthetic images were of two kinds:

- An image of a circle circumscribed by a square which was corrupted by varying amounts of additive Gaussian noise. This image was used to evaluate the ability of the edge-labeling algorithms to detect edges of varying orientation and to find the corners of the square which represent points of high curvature.

- In addition to being subject to thermal noise contamination, realistic edges are sampled by imaging devices of finite resolution and are subject to illumination gradients. As a consequence, the sampled edges of physical objects in natural scenes will not present step-profiles in the image luminance function. To assess the effectiveness of relaxation labeling to correctly identify the position of more realistic edge-types we have generated sigmoid edge-profiles of varying width. This has been achieved by applying low-pass filters of varying mask size to a ramp structure of a few pixels width. Thermal noise has subsequently been added to the profiles.

In addition to these synthetic images, the edge-labeling algorithms have been applied to natural images. In particular, we have studied images of cluttered scenes where realistic edges do not necessarily have a step-profile and noise may be present due to the image sampling procedure.

C. Results

Fig. 4 shows the results of successive iterations of the dictionary-based labeling algorithm applied to the circle-inside-square image in which the signal-to-noise ratio (SNR) is 25:6; the intensities of the pixels are proportional to the probability of the non-edge-label. After eight iterations of the evidence-combining formula, the process has essentially converged and there are very few pixels

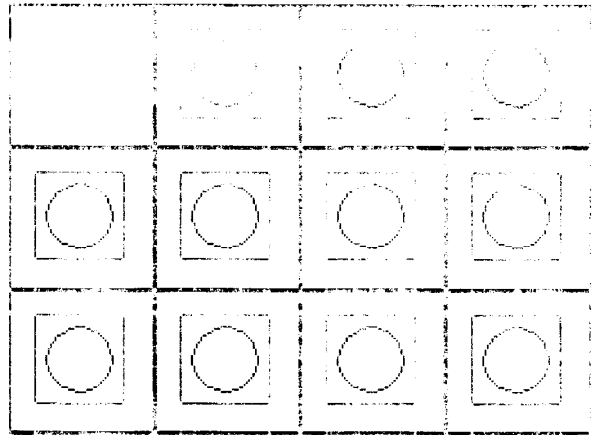


Fig. 4. Successive iterations of dictionary-based relaxation.

for which the probability is neither zero nor unity. It can be seen there are no inconsistent labelings present in the scene and all edges are one pixel thick. It is also worth noting that all four corners of the square are well reconstructed.

In order to evaluate the relative performance of the edge-labeling algorithms described in Section IV-A, a series of circle-inside-square images with signal-to-noise ratios of $\frac{25}{2}$, $\frac{25}{4}$, $\frac{25}{6}$, $\frac{25}{8}$, and $\frac{25}{10}$ were used. Fig. 5 shows a comparison of results obtained using dictionary-driven relaxation, Hilditch edge-thinning, and Spacek's algorithm without filtering applied to these images. At the lowest level of noise all four algorithms perform well. When the signal-to-noise ratio falls below $\frac{25}{4}$, Hilditch edge-thinning fails. When the signal-to-noise ratio is less than $\frac{25}{6}$, Spacek's edge-labeling algorithm is no longer able to label genuine edge-pixels without also labeling a large number of noise pixels. The dictionary-based relaxation approach gives good results for values of the signal-to-noise ratio down to $\frac{25}{8}$. Even when the signal-to-noise ratio is below this value, the dictionary-based approach labels a reasonable fraction of genuine edge-pixels without labeling an unacceptably large number of non-edge-pixels. Another encouraging feature of the edge-segments reconstructed by the dictionary-based approach is their strong connectivity.

The dictionary-based approach can also be used to label the output from filters which suppress the noise and produce multiresolution representations of image data; Spacek's circularly symmetric operator is such a filter. When used to suppress noise, these filters have the property of smoothing the high-frequency-spatial component in an image. Under certain circumstances this may be an undesirable property since genuine edge-features such as corners or fine details are removed. Since dictionary-based relaxation models the noise generation process and can reliably reconstruct edge-segments at higher levels of noise than other labeling algorithms, it may be used to label edges using a smaller mask size than, for instance,

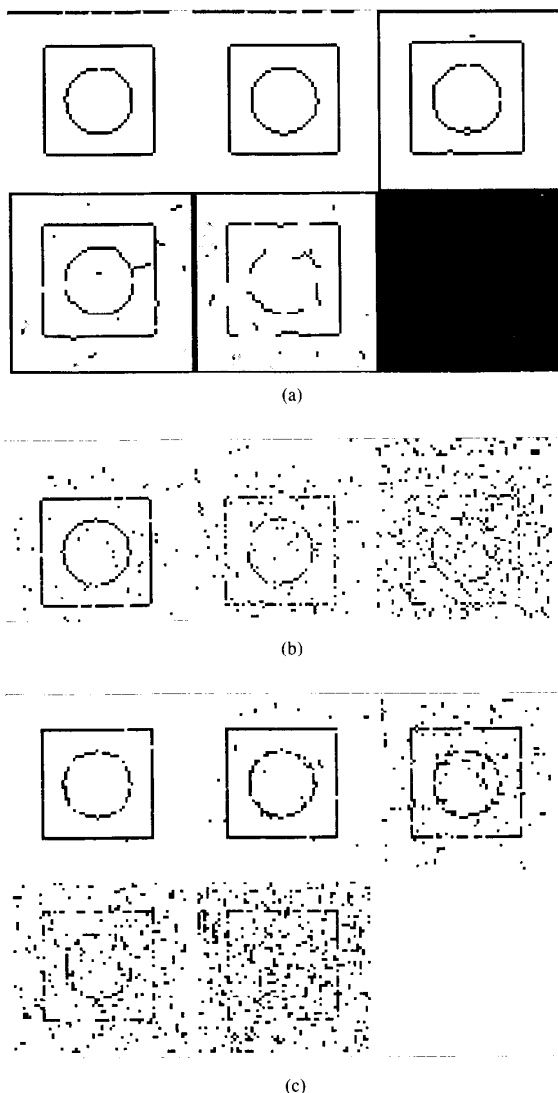


Fig. 5. (a) Dictionary-based relaxation applied to a synthetic image with different levels of additive Gaussian noise; i) SNR = 25/2, ii) SNR = 25/4, iii) SNR = 25/6, iv) SNR = 25/8, v) SNR = 25/10. (b) Hil-ditch edge-thinning applied to a synthetic image with different levels of additive Gaussian noise; i) SNR = 25/2, ii) SNR = 25/4, iii) SNR = 25/6. (c) Spacek's labeling procedure applied to a synthetic image with different levels of additive Gaussian noise; i) SNR = 25/2, ii) SNR = 25/4, iii) SNR = 25/6, iv) SNR = 25/8, v) SNR = 25/10.

Spacek's polynomial method. In order to demonstrate this effect a circle-inside-square image with signal-to-noise ratio of 25:15 has been convolved with circularly symmetric Spacek filters of mask size (w) 3, 5, 7, 9, and 11 pixels. The dictionary-based labeling of the filtered image successfully reconstructs edge-segments for a mask size of 5 pixels, while Spacek's polynomial method only succeeds if the mask size exceeds 9 pixels; these results are demonstrated in Fig. 6. When the mask size used exceeds 9 pixels the relaxation approach begins to label apparently

spurious edges in the proximity of the boundaries of the figures in the synthetic image. This problem appears to be the consequence of the dilation of the intensity gradient by the large filter mask and may be rectified by subsampling the filtered image.

Fig. 7 is a comparison of the dictionary-based approach and Canny's postprocessing applied to a synthetic circle-inside-square image of signal-to-noise ratio of 25:8. Fig. 7(a) shows the best result that could be obtained when Canny's method was applied to the raw image; no filtering has been employed. The result is poor in comparison with that shown in Fig. 7(b) which was obtained using the dictionary-based approach. A much better result is obtained when Canny's postprocessing is applied the output of a Gaussian filter. Fig. 7(c) shows the best labeling that could be obtained after filtering with a support mask of 5×5 had been performed. Filtering based on a smaller support mask fails to adequately label the physical edges of the figures in the image. The erosion of corners is very evident in Fig. 7(c). Finally, Fig. 7(d) shows the result of applying dictionary-based relaxation to the output of a Gaussian filter; this is the most satisfactory of the four labelings.

From this experimentation with Canny's postprocessing, it is clear that the method only performs well when applied to a filtered representation of the image. It cannot cope with raw noise. Although the postprocessing has been applied to the output of a Gaussian filter, it could equally well have been applied to the output of Spacek's optimal filter.

The results described above indicate that dictionary-based relaxation labeling can be successfully applied to images containing step edges which have been corrupted by additive noise. However, the model used to calculate the initial *a posteriori* probabilities does not depend on the profile of the edge. In order to investigate the behavior of the relaxation approach when the gray-scale profile departs radically from that of a step edge, we have studied the behavior of the labeling algorithms when applied to synthetic profiles.

We have generated a sigmoid profile of width 2 pixels. This edge-profile was corrupted with additive Gaussian noise with standard deviation varying from zero gray-levels at the top of the image to 32 gray-levels at the bottom of the image, i.e., the signal-to-noise ratio has a minimum value of 4. The corrupted image was then processed using filters of mask size two and four pixels. The resulting image simulates the combined effects of sensor noise and finite sampling width. Without any filtering of the gray-scale intensities the relaxation approach is able to reconstruct the edge provided that the signal-to-noise ratio is not less than 6.4; after filtering with mask size 4 pixels, the relaxation approach successfully reconstructs the entire edge at all levels of noise. These results are encouraging since they indicate that the relaxation approach is both robust to noise and the width of the profile. These results are demonstrated in Fig. 8.

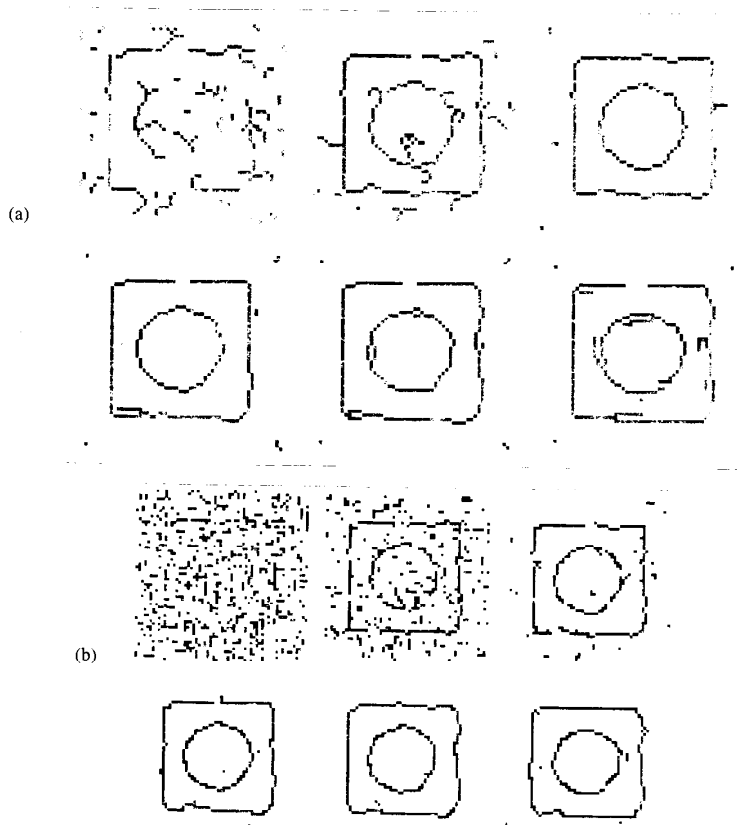


Fig. 6. (a) Dictionary-based relaxation applied to a synthetic image which has been convolved with circularly symmetric filters of different mask size; i) $w = 1$, ii) $w = 3$, iii) $w = 5$, iv) $w = 7$, v) $w = 9$, vi) $w = 11$. (b) Spacek's edge-operator with filters of different mask size applied to a synthetic image; i) $w = 1$, ii) $w = 3$, iii) $w = 5$, iv) $w = 7$, v) $w = 9$, vi) $w = 11$.

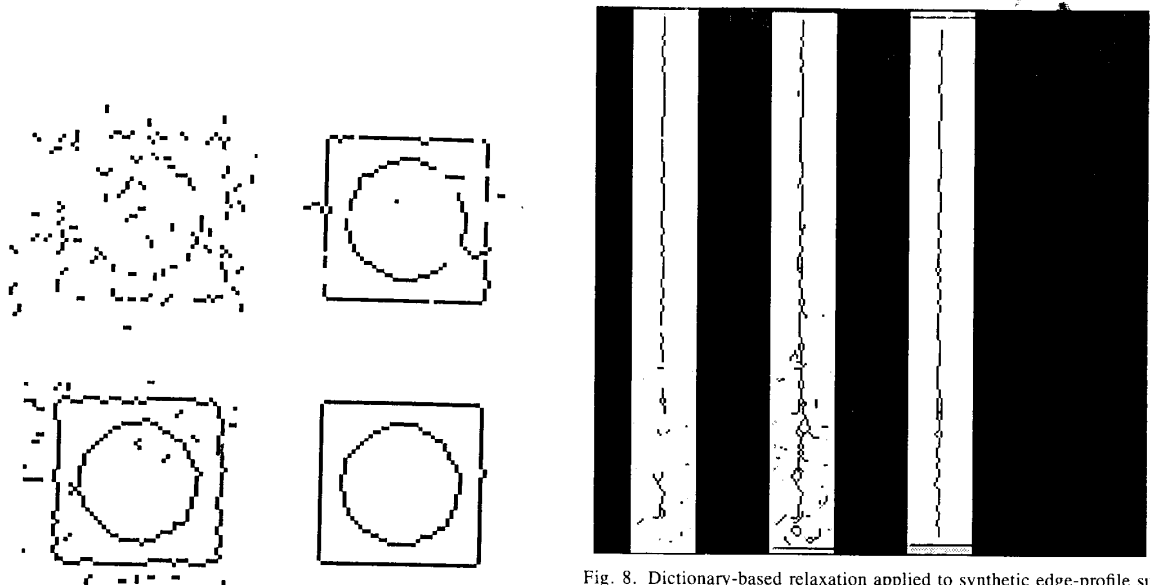


Fig. 7. Comparison of dictionary-based relaxation and Canny's algorithm for an SNR = 25:8 image.

Fig. 8. Dictionary-based relaxation applied to synthetic edge-profile subject to varying amounts of additive Gaussian noise and smoothed using low-pass filters of different width: i) $w = 1$, ii) $w = 2$, iii) $w = 4$. The signal-to-noise ratio varies from 128/0 at the top of the image to 128/32 at the bottom.

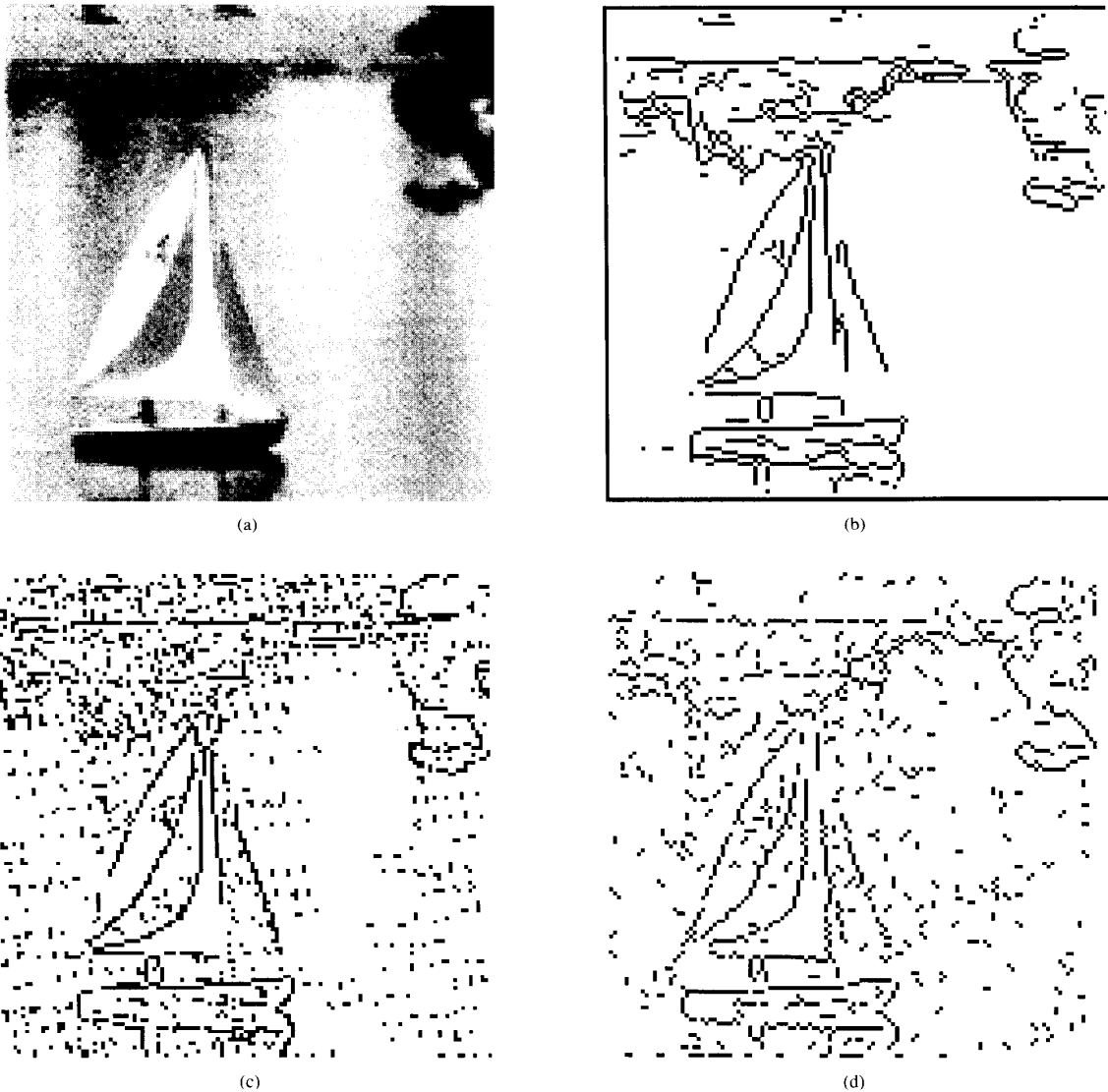


Fig. 9. (a) Original image of a natural scene. (b) Dictionary-based relaxation applied to an image of a natural scene. (c) Spacek's second-order surface model applied to an image of a natural scene. (d) Canny's post-processing applied to an image of a natural scene.

We have performed extensive comparative experiments on images of natural scenes. Figs. 9, 10, and 11 show examples of the comparative performance of dictionary-based relaxation and the postprocessing algorithms of Spacek and Canny. The image used for the comparison in Fig. 9 was chosen because of its noisy and textured nature. Both Spacek and Canny prove susceptible to the noise component and cannot cleanly segment out the meaningful edge features. Dictionary-based relaxation, on the other hand, is capable of locating the important edge-information without labeling an unacceptable amount of extraneous edge-structure. Fig. 10 shows a comparison of

the three techniques for a less noisy and uncluttered scene. All three techniques perform well, however, the edges located by dictionary-based relaxation display better connectivity properties. Finally, Fig. 11 shows the comparison on a fairly complex scene. In the foreground there are isolated edges corresponding to shape boundaries, while the background is dense with occluding edges. All three approaches appear to label successfully a large number of genuine edges. However, the relaxation labeling algorithm reconstructs edges which have a much stronger tendency to be connected and finds fewer edge-features which appear to be artifacts of texture or noise.

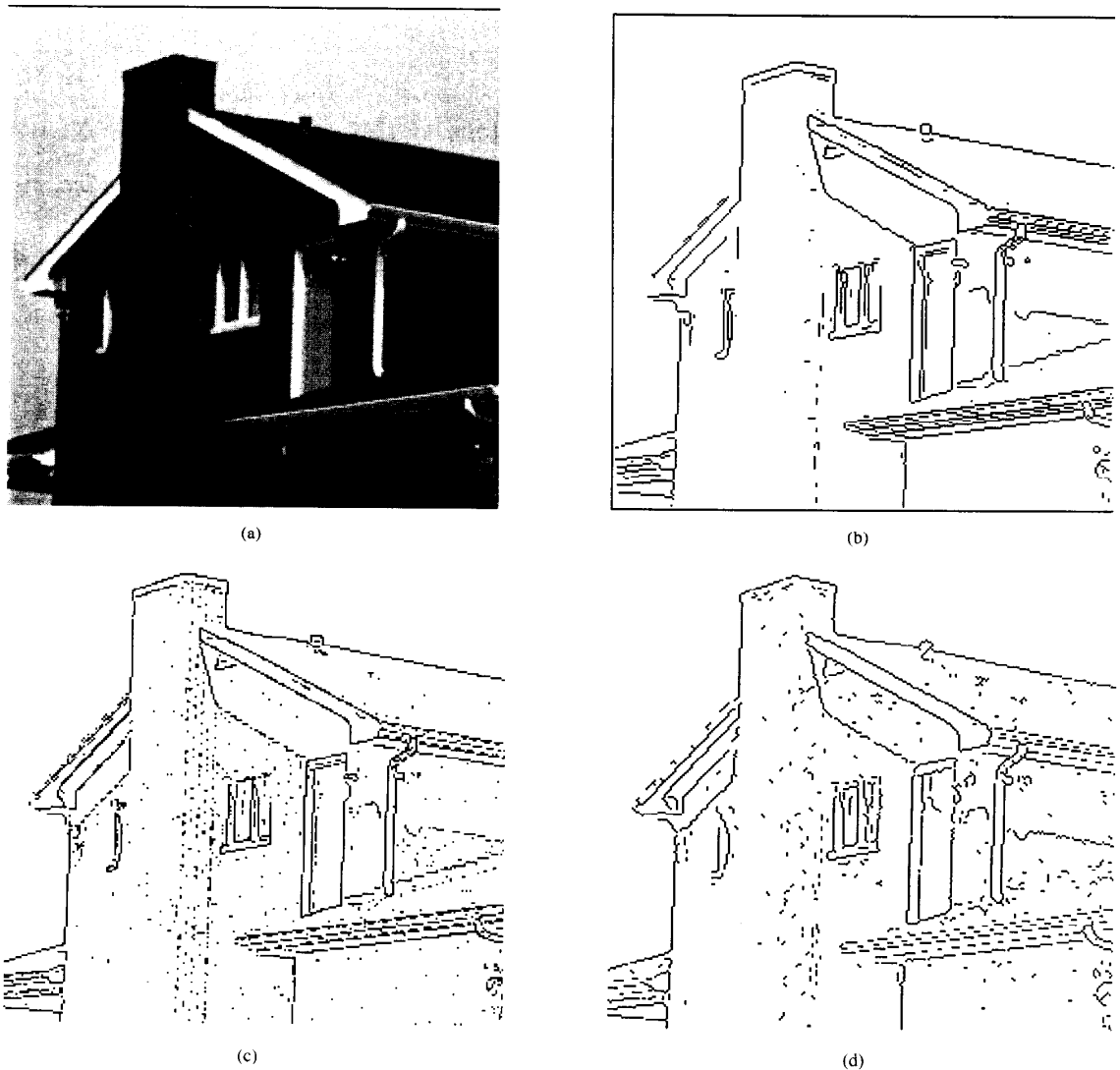


Fig. 10. (a) Original image of a natural scene. (b) Dictionary-based relaxation applied to an image of a natural scene. (c) Spacek's second-order surface model applied to an image of a natural scene. (d) Canny's post-processing applied to an image of a natural scene.

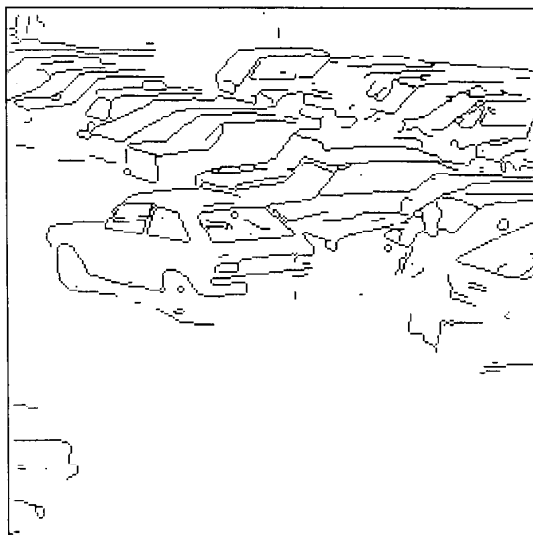
In order to demonstrate that dictionary-based relaxation can succeed in labeling a natural variety of edge-types in the presence of noise, we have corrupted the image used for the experimentation shown in Fig. 11 with additive Gaussian noise of standard deviation 15 gray-levels. Fig. 12(a) shows the result of applying dictionary-based relaxation to the corrupted image, Fig. 12(b) shows the result of applying Spacek's labeling algorithm to the unfiltered image, and Fig. 12(c) shows the result of applying Spacek's labeling algorithm when the image has been filtered with a mask of size 5 pixels. The labeling produced by Spacek's algorithm without a filter operation is clearly inferior to dictionary-based relaxation; it is only when a fil-

ter mask of 5 pixels is applied that the two results achieve comparability.

In practical applications of edge-labeling an important consideration is automatic adaptation to varying noise conditions. As indicated above, in the dictionary-based approach the response to differing levels of noise is controlled by the parameter σ used in the calculation of P_{residual} while in Spacek's strategy it is controlled by the mask size used in the filtering operation. One strategy for adaptively calculating σ is to adopt an estimation procedure similar to that suggested in [15]. In the studies reported here, the value of σ that gave the best subjective performance was greater than the additive noise compo-



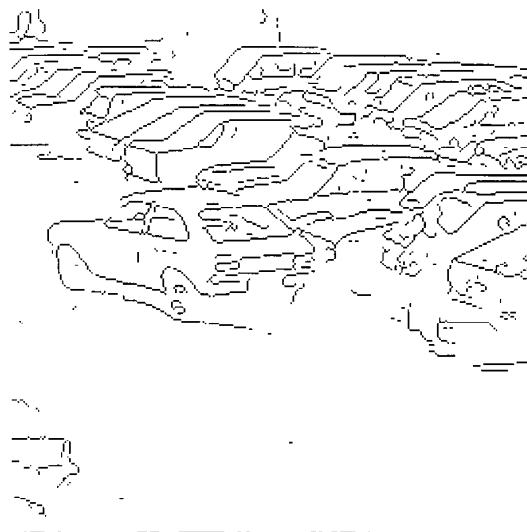
(a)



(b)



(c)



(d)

Fig. 11. (a) Original image of a natural scene. (b) Dictionary-based relaxation applied to an image of a natural scene. (c) Spacek's second-order surface model applied to an image of a natural scene. (d) Canny's post-processing applied to an image of a natural scene.

nent in the images; this is reconcilable with the fact that no attempt has been made at accounting for quantization errors.

V. CONCLUSIONS

We have demonstrated that probabilistic relaxation can be successfully applied to the problem of edge-labeling. The success of the application derives from two features of the way in which the relaxation scheme is formulated. Firstly, we commence with a specification of the proba-

bilistic framework that is used to represent the world model; this ensures internal consistency. Secondly, we represent prior knowledge of the structure of the application using a dictionary of labeling possibilities for the entire context-conveying neighborhood for each object; this is to be contrasted with the use of factorizations of the neighborhood which involve limiting approximations and reduce the representational capacity of the scheme. The resulting algorithm combines evidence for the presence of edges by drawing on prior knowledge of the permitted label structures.

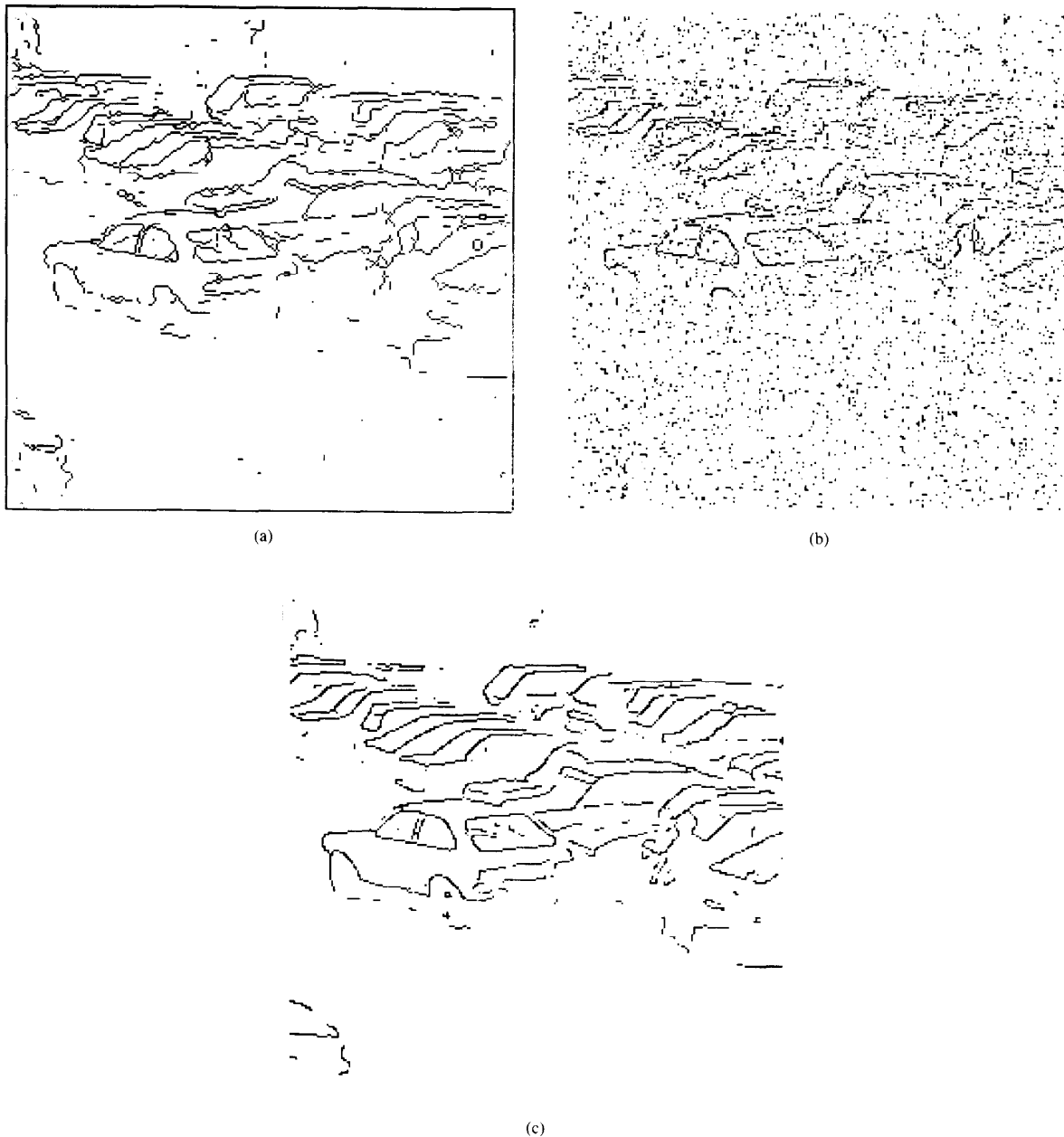


Fig. 12. (a) Dictionary-based relaxation applied to an image of a natural scene corrupted by additive Gaussian noise of standard deviation 15 gray-levels. (b) Spacek's labeling algorithm applied to an image of a natural scene corrupted by additive Gaussian noise of standard deviation 15 gray-levels. (c) Spacek's labeling algorithm with filter mask size 5 pixels applied to a natural scene corrupted by additive Gaussian noise of standard deviation 15 gray-levels.

The consequence of the improved formulation is a computationally efficient algorithm which is comparable in performance to edge-detection strategies which are acknowledged as being successful. In particular, the dictionary-based approach is capable of enhancing edge-structures

in the presence of considerable noise without the need to apply filters which involve a large support mask. This has the advantage that high-frequency edge-features such as corners are not sacrificed at the expense of noise suppression.

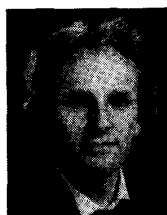
REFERENCES

- [1] A. Rosenfeld, R. A. Hummel, and S. W. Zucker, "Scene labeling by relaxation operations," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-6, pp. 420-433, 1976.
- [2] J. Kittler and J. Illingworth, "A review of relaxation labeling algorithms," *Image Vision Comput.*, vol. 3, pp. 206-216, 1985.
- [3] J. Kittler and E. R. Hancock, "Combining evidence in probabilistic relaxation," *Int. J. Pattern Recognition Artificial Intell.*, vol. 3, pp. 29-52, 1989.
- [4] O. D. Faugeras and M. Berthod, "Improving consistency and reducing ambiguity in stochastic labeling: An optimization approach," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-3, pp. 412-424, 1981.
- [5] R. A. Hummel and S. W. Zucker, "On the foundations of relaxation labeling processes," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-5, pp. 267-287, 1983.
- [6] J. Kittler and E. R. Hancock, "Contextual decision rule for region analysis," *Image Vision Comput.*, vol. 5, pp. 145-155, 1987.
- [7] E. R. Hancock and J. Kittler, "A list-driven contextual decision rule," in *Proc. 5th Scandinavian Conf. Image Analysis*, vol. 2, 1987, pp. 555-562.
- [8] S. W. Zucker, R. A. Hummel, and A. Rosenfeld, "An application of relaxation labeling to line and curve enhancement," *IEEE Trans. Comput.*, vol. C-26, pp. 394-403, 1977.
- [9] L. A. Spacek, "Edge-detection and motion detection," *Image Vision Comput.*, vol. 4, pp. 43-56, 1986.
- [10] S. Peleg and A. Rosenfeld, "Determining compatibility coefficients for curve enhancement relaxation processes," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-8, pp. 548-555, 1978.
- [11] J. Hilditch, "Linear skeletons form square cupboards," *Machine Intell.*, vol. 6, pp. 403-420, 1969.
- [12] J. Kittler, "Compatibility and support functions in probabilistic relaxation," in *Proc. 8th ICPR*, Paris, 1986.
- [13] J. Kittler and J. Föglein, "On compatibility and support functions in probabilistic relaxation," *Comput. Vision, Graphics, Image Processing*, 1986.
- [14] J. Kittler, "Relaxation labeling," in *Pattern Recognition Theory and Practice*, P. A. Devijver and J. Kittler, Eds. Hiedelberg: Springer-Verlag, 1987.
- [15] J. Kittler, J. Illingworth, J. Föglein, and K. Paler, "An automatic thresholding algorithm and its performance," in *Proc. 7th ICPR*, Montreal, 1984.
- [16] J. F. Canny, "Finding edges and lines in images," M.I.T. Tech. Rep. 720, 1983.
- [17] J. F. Canny, "A computational approach to edge-detection," *IEEE Trans. Pattern. Anal. Machine Intell.*, vol. PAMI-8, pp. 679-700, 1986.
- [18] V. Nalwa, "On detecting edges," *IEEE Trans. Pattern. Anal. Machine Intell.*, vol. PAMI-8, pp. 701-714, 1986.
- [19] D. Marr and E. Hildreth, "Theory of edge-detection," in *Proc. Roy. Soc. London*, vol. B207, pp. 187-212, 1980.
- [20] R. M. Haralick and L. T. Watson, "A facet model for image data," *Comput. Vision, Graphics, Image Processing*, vol. 15, pp. 113-129, 1981.
- [21] D. L. Waltz, "Understanding line drawings of scenes with shadows," in *The Psychology of Computer Vision*, P. H. Winston, Ed. New York: McGraw-Hill, 1975.
- [22] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and Bayesian restoration of images," *IEEE Trans. Pattern. Anal. Machine Intell.*, vol. PAMI-6, pp. 721-741, 1984.
- [23] H. Derin, H. Elliot, R. Cristi, and D. Geman, "Bayes smoothing algorithms for segmentation of binary images modeled by Markov random fields," *IEEE Trans. Pattern. Anal. Machine Intell.*, vol. PAMI-6, pp. 707-720, 1984.
- [24] F. R. Hansen and H. Elliot, "Image segmentation using simple Markov random fields," *Comput. Vision, Graphics, Image Processing*, vol. 20, pp. 101-132, 1982.
- [25] R. M. Haralick and J. S. J. Lee, "Context dependent edge-detection," in *Proc. 9th ICPR*, Rome, vol. 1, 1989, pp. 203-207.
- [26] J. E. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Stat. Soc.*, series B, to appear.



Edwin R. Hancock received the B.Sc. degree in physics and the Ph.D. degree in high energy physics from the University of Durham, UK, in 1977 and 1981, respectively.

In 1981 he joined the Rutherford Appleton Laboratory to work on an experiment to measure charmed particle lifetimes using high resolution imaging techniques at the Stanford Linear Accelerator Center. Since 1985, he has been doing research in the areas of pattern recognition and computer vision, his current interests being in the methodology and the application of relaxation processes to image interpretation problems. He is also an Associate Lecturer in the Department of Electrical Engineering at the University of Surrey. He has published over 25 papers in refereed journals and conferences.



Josef Kittler (M'74) has been working in the fields of pattern recognition, image analysis, and computer vision since 1970. He is the author or coauthor of more than 100 publications which include 5 books (one coauthored and four edited) and more than 40 papers in refereed international journals.

Dr. Kittler is a member of the editorial boards of IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, *Pattern Recognition Journal*, *Image and Vision Computing*, *Pattern*

Recognition Letters, and the *International Journal of Pattern Recognition and Artificial Intelligence*.