



## The 2.1-D Sketch

The Harvard community has made this article openly available. [Please share](#) how this access benefits you. Your story matters

Citation	Nitzberg, Mark, and David Bryant Mumford. 1990. The 2.1-D sketch. In Proceedings of the Third International Conference on Computer Vision: December 4 - 7, 1990, Osaka, Japan, ed. IEEE Computer Society, 138-144. Los Alamitos, CA: IEEE Computer Society Press.
Published Version	doi:10.1109/ICCV.1990.139511
Citable link	<a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:3637122">http://nrs.harvard.edu/urn-3:HUL.InstRepos:3637122</a>
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a>

# The 2.1-D Sketch\*

Mark Nitzberg

David Mumford

Division of Applied Sciences  
Harvard University  
Cambridge, MA 02138 USA

*e-mail address:* mumford@math.harvard.edu

## Abstract

*This paper describes a model for image segmentation that tries to capture the the low-level depth reconstruction exhibited in early human vision, giving an important role to edge terminations.*

*The problem is to find a decomposition of the domain  $D$  of an image that has a minimum of disrupted edges—junctions of edges, crack tips, corners, and cusps—by creating suitable continuations for the disrupted edges behind occluding regions. The result is a decomposition of  $D$  into overlapping regions  $R_1 \cup \dots \cup R_n$  ordered by occlusion, which we call the 2.1-D Sketch.*

*Expressed as a minimization problem, the model gives rise to a family of optimal contours, called nonlinear splines, that minimize length and the square of curvature. These are essential in the construction of the 2.1-D sketch of an image, as the continuations of disrupted edges.*

*The paper describes an algorithm that constructs the 2.1-D sketch of an image, and gives results for several example images. The algorithm yields the same interpretations of optical illusions as the human visual system.*

## Introduction

One of the principal goals of low-level vision is to *segment* an image. Thus if  $g(x, y)$ ,  $(x, y) \in D$ , represents an observed image (i.e. the light intensity striking the lens from direction  $(x, y)$ ), then we want to segment the domain  $D$ , i.e. partition  $D$  into regions  $R_1, \dots, R_k$  such that  $R_i$  is the part of the image in which the nearest object is some object  $O_i$  and on the boundary between

any two regions  $R_i$  and  $R_j$ , object  $O_i$  occludes object  $O_j$  or vice-versa. Another of the goals of low-level vision is to compute or estimate what David Marr called the  $2\frac{1}{2}$ -D sketch associated to the image [1], i.e. the depth image  $z(x, y)$  recording the distance from the lens to the nearest object in the direction  $(x, y)$  and its normalized partial derivatives:

$$p(x, y) = z_x / \sqrt{1 + z_x^2 + z_y^2}$$

$$q(x, y) = z_y / \sqrt{1 + z_x^2 + z_y^2}$$

Marr proposed multiple sources of information contained in the intensity image  $g(x, y)$  from which one could hope to estimate the  $2\frac{1}{2}$ -D sketch  $(z, p, q)$ , but it has proved hard to implement his program except in cases where very accurate stereo or motion data is available, or where the lighting and surface reflectances are heavily constrained. In this paper, we want to propose a synthesis of these two goals, segmentation and the  $2\frac{1}{2}$ -D sketch, which greatly simplifies the numerical burden of the  $2\frac{1}{2}$ -D sketch and at the same time simplifies 2D segmentation by incorporating occlusion explicitly. We will call this the 2.1-D sketch.

Consider figure 1(a), an image of several blades of grass against a blurred, distant background. A portion of this image, somewhat simplified, is shown segmented in figure 1(b) using 9 disjoint regions. However, the 9 regions do not correspond to 9 distinct objects in the world: there are only 4 objects reflecting light - the 3 blades of grass and the background "object". Although each of these objects lies at varying depth, there is a simple ordering of the objects telling you which objects occlude which others. So we can describe the scene by a stage set with 4 "wings", transparent except where they contain an object (the background is last and is everywhere opaque). This is shown in figure 1(c).

\*supported by US Army Research Office grant DAAL03-86-K-0171, and by a grant from the MacArthur Foundation

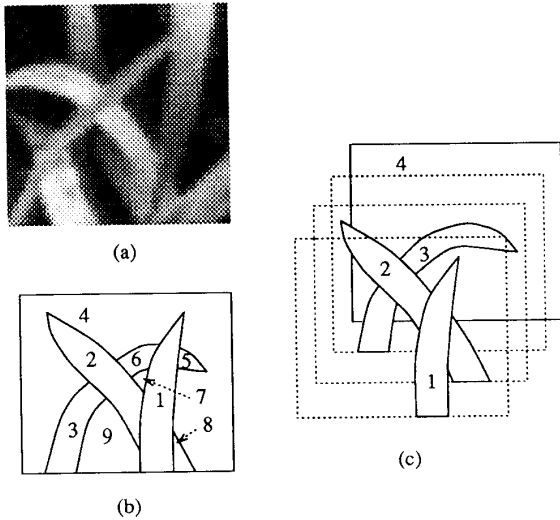


Figure 1: (a) Blades of grass image, (b) a disjoint segmentation; (c) the 2.1-D Sketch

This is what we mean by a 2.1-D sketch: it is a set of regions  $R_i$  in the domain  $D$  of the image which fill up  $D$  but which may overlap, in general, plus a partial ordering  $>$  on the regions indicating which are in front of which others. Often there will be a background object  $R_0$  behind all others for which  $R_0 = D$ . Our contention is that this type of segmentation is more natural than the kind with disjoint, unordered  $R_i$ ; and that it captures the most accessible part of the  $2\frac{1}{2}$ -D sketch.

The chief reason for which we expect the 2.1-D sketch to be readily computable is the presence of T-junctions. T-junctions are points where the edges in the image form a "T", one edge  $\Gamma_1$  ending abruptly in the middle of a second edge  $\Gamma_2$ . Such points nearly always arise because  $\Gamma_2$  is an occlusion edge and  $\Gamma_1$  is any kind of edge (occlusion edge, shadow edge, surface marking edge) on a more distant object whose continuation disappears behind  $\Gamma_2$ . Our experience with a variety of photographs in many different settings is that there are very often enough T-junctions to order most of the visible objects in depth. Shading heuristics (not precise shape-from-shading calculations) often provide important complementary clues.

The importance of T-junctions in the human visual system has been known for a long time, but its role and power has been greatly clarified by recent work. In particular, it has become increasingly clear that T-junctions are computed early in the visual process and are not merely part of an object recognition paradigm

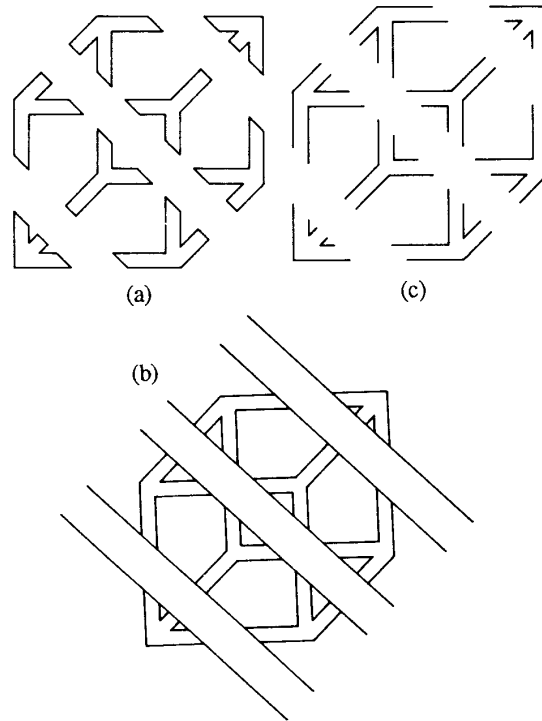


Figure 2: A demonstration from Kanizsa of the importance of T-junctions and terminators.

as in the early blocks-world algorithms of Guzman, Roberts, Waltz, etc. The gestalt school of psychology and, particularly, the contemporary psychologist Gaetano Kanizsa has made a thorough and deep analysis of T-junctions [2]. Consider figure 2 from Kanizsa. 2(a) and 2(b) differ only in the addition of diagonal lines which change the corners in 2(a) to T-junctions in 2(b); 2(b) is unmistakably 3-dimensional. 2(a) and 2(c) differ only in the subtraction of short connecting lines which change corners in 2(a) to terminators in 2(c).

We see that a terminating line is a weak form of T-junction and that several aligned terminators are a strong clue to occlusion. Likewise corners, especially when pairs of their edges are aligned, are degenerate forms of T-junctions, as in Kanizsa's pac-man illusion, see figure 3(a).

In all these cases, the mind seems to create a 3D scene in which occluded parts of visible objects are reconstructed. In the pac-man case, the mind goes further and creates missing outlines of the nearer occluding triangle, explaining their absence in the raw data by an accidental match of the lightness of the occluding object and the far background.

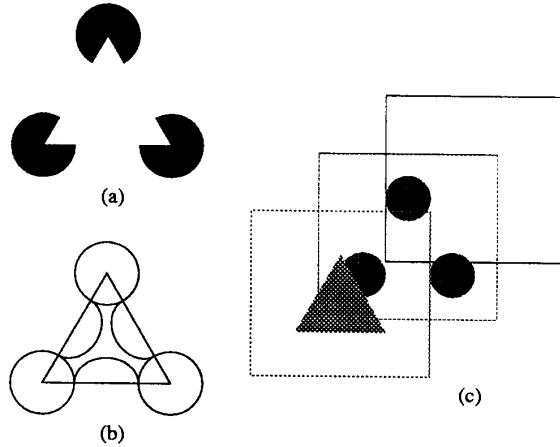


Figure 3: (a) Kanizsa's pac-man illusion; (b) a disjoint segmentation with edge completions; (c) the 2.1-D Sketch

A striking confirmation of the reality of these so-called illusions and the illusory contours that we feel we see was found by R. von der Heydt and his lab [3]. The responses of single cells in visual areas are codified by describing their visual field: the area within which moving or stationary bars and edges produces activity. They found, however, that many cells in visual area V2<sup>1</sup> responded when no actual stimulus was present in their visual field, but when edges outside this field produce illusory contours that cross this field. Finally, we want to mention the beautiful experiments of K. Nakayama who has explored human perception of data in which various clues for depth-T-junctions, stereo, motion, etc.-conflict [4]. His results show in many cases that T-junctions override other clues and are always powerful organizing forces in an image.

## Energy functionals

The piece-wise smooth model of the segmentation problem in computer vision asks how to clip a picture into as few and simple pieces as possible while keeping the color of each piece as smooth and/or slowly varying as possible. One approach to the problem, taken by Mumford and Shah [5], is to define a func-

<sup>1</sup>Known as Brodmann area 18 in man, this area is adjacent to the primary visual area V1 (= area 17 = striate cortex) and is a recipient of a high proportion of its axonal output.

tional that takes its minimum at an optimal piece-wise smooth approximation to a given image. The image is a function  $g$  defined on a domain  $D$  in the plane. It is approximated by a function  $f$ , which is smooth except at a finite set  $\Gamma$  of piece-wise  $C^1$  contours which meet  $\partial D$  and meet each other only at their endpoints. The functional defined below gives a measure of the match between an image  $g$  and a segmentation  $f, \Gamma$ :

$$E_{M.S}(f, \Gamma) = \mu^2 \int_D (f-g)^2 dx + \int_{D \setminus \Gamma} \|\nabla f\|^2 dx + \nu \int_{\Gamma} ds.$$

The first term asks that  $f$  approximate  $g$ , the second asks that  $f$  vary slowly except at boundaries, and the third asks that the set of contours be as short, and hence as simple and straight as possible. The contours of  $\Gamma$  cut  $D$  into a finite set of disjoint regions  $R_1, \dots, R_n$ , the connected components of  $D \setminus \Gamma$ .

In this paper, however, we seek a model that incorporates partially the way that  $g$  derives from a 2-D projection of a 3-D scene. Rather than base our 2.1-D model on a set of curves  $\Gamma$  that cuts  $D$  into disjoint regions, we ask for a set of regions  $R_i$  whose union equals  $D$ , and with a partial ordering that represents relative depth. The overlapping of regions gives in a sense the most primitive depth information. The domain  $D$  is considered as a window that reveals the value of  $g$  only on a portion of the plane. As a result, contour integrals will exclude portions of a contour that coincide with the boundary of  $D$ .

Here is a variational formulation of the problem of finding the 2.1-D sketch. We seek a functional  $E_{2.1}$  much like  $E_{M.S}$  that achieves a minimum at the optimal overlapping segmentation of  $g$ . Let  $\{R_1, \dots, R_n\}$  be a set of regions such that  $\bigcup_i R_i = D$ , with a partial ordering  $<$  that represents occlusion.

$$R'_i = R_i \setminus \bigcup_{R_j < R_i} R_j$$

is the "visible" portion of  $R_i$ . Throughout the paper,  $R_i$  denotes a closed subset of  $D$  with piecewise smooth boundary and connected interior. The expression  $(\{R_i\}, <)$  denotes an ordered set of overlapping regions, which we will call a *segmentation*.

We then define the energy  $E_{2.1}(\{R_i\}, <)$  as

$$\sum_{i=1}^n \left( \mu^2 \int_{R'_i} (g - m_i)^2 dx + \epsilon \int_{R_i} dx + \int_{\partial R_i \setminus \partial D} \phi(\kappa) ds \right).$$

In this formula,  $m_i$  is the mean of  $g$  on  $R'_i$ , and  $\kappa$  is the curvature of  $\partial R_i$ , i.e.  $\ddot{\gamma}$  where  $\gamma$  parameterizes  $\partial R_i$  by arc length. The function  $\phi : \mathbf{R} \rightarrow \mathbf{R}$  is defined by

$$\phi(\kappa) = \begin{cases} \nu + \alpha \kappa^2 & \text{for } |\kappa| < \beta/\alpha \\ \nu + \beta |\kappa| & \text{for } |\kappa| \geq \beta/\alpha \end{cases}$$

The scalar constants  $\mu$ ,  $\nu$ ,  $\epsilon$ ,  $\alpha$  and  $\beta$  in the definition of  $\phi$ , determine the characteristics of a segmentation which minimizes  $E_{2.1}$ . Their dimensions are:

$$\begin{aligned} \mu &\sim \text{intensity}^{-1} \cdot \text{dist.}^{-1} \\ \nu &\sim \text{dist.}^{-1} \\ \alpha &\sim \text{dist.} \\ \beta &\sim \text{dimensionless} \\ \epsilon &\sim \text{dist.}^{-2} \end{aligned}$$

This simple choice of model applies to a variety of pictures. Its simplicity necessarily restricts the kinds of pictures for which it is effective.

Firstly, the model does not allow self-overlapping “woven” regions such as a garden hose would project, nor folded regions such as produced by an image of a sleeve whose edge disappears around the back of an arm. This can be solved with a model that assigns local occlusion relations along edges rather than trying to find a global organization.

Secondly, the model is piece-wise constant: each region  $R_i$  is given a constant intensity  $m_i$ . This means that the simple curved surface of a coffee mug, which can project a single region with greatly varying brightness, may wrongly be divided into several regions in a minimum of  $E_{2.1}$ . It also means that crack tips never appear in any  $R_i$ . This is solved by replacing the constant  $m_i$  with a smooth approximating function  $f_i$  on  $R_i$ , and adding a penalty term  $\sum_i \int_{R_i} \|\nabla f_i\|^2$  to ask that  $f_i$  vary smoothly on  $R_i$ .

Thirdly, transparency and shadows are not treated. This can be solved by giving nearer regions the choice of either *occluding* farther regions, or *modifying* them, i.e., shadows darken farther regions, while water distorts what is behind it.

## Properties of $E$

The first and second terms of  $E_{2.1}$  are easily seen to give, on  $R'_i$ :

$$(\text{variance}_{R'_i}(g) + \epsilon) \cdot \text{area}(R'_i).$$

This is the basic term that keeps  $g$  close to constant on each region; without it, the trivial segmentation ( $\{D\}$ ) achieves a minimum of  $E$ .

An interesting variant, due to Y. Leclerc, incorporating these terms, is

$$\log(\text{variance}_{R'_i}(g) + \epsilon) \cdot \text{area}(R'_i)$$

The term then approaches  $\log(\epsilon) \cdot \text{area}(R'_i)$ , as  $\text{variance}_{R'_i}(g) \rightarrow 0$ . This variant is a first step to incorporating texture boundaries into energy functionals:

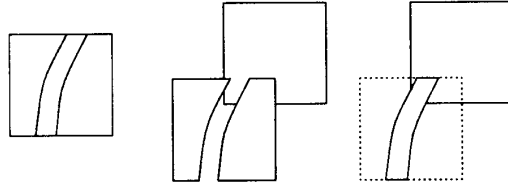


Figure 4: A simple image with two edges, and two possible 2.1-D Sketches.

Leclerc derived it by a minimum description length argument, employing both the mean and the variance of the intensity on each region. It causes a segmentation of regions with the same mean intensity but sufficiently different variances.

The second term of  $E_{2.1}$  gives a penalty for the total area of all regions. Its purpose is to encourage segmentations in which small regions occlude larger ones. Without an area term, the segmentations shown for figure 4(a), in figures 4(b) and (c) give the same value for  $E$ ; with an area term (c) gives a lower value for  $E$ . The parameter  $\epsilon$  should be very small, because the area term is intended only as a tie-breaker in otherwise ambiguous situations.

The third term asks that the boundaries of regions in the interior of  $D$  be short and not too curvy. It equals  $\nu$  times the length of  $\partial R_i - \partial D$  plus  $\alpha$  times the integral of the curvature squared along the same boundary, except where curvature exceeds  $\beta/\alpha$ , whereupon the term becomes  $\beta$  times the integral of the curvature. Note that the integral of the curvature over a piece  $\gamma$  of  $\partial R_i$  is just the total angle through which  $\gamma$  turns. This term discourages region boundaries from tracing circuitous level-curves in their effort to minimize the variance term. Moreover, this term determines a family of contours that comprises the ideal continuations of disrupted edges behind occluding regions, discussed in the next section.

The purpose of  $\phi(\cdot)$  changing from a quadratic to a linear function at  $\beta/\alpha$  is to allow us to extend the definition of the second term to a contour which is smooth except at a finite number of infinite-curvature points, or “corners.” Suppose a contour  $\gamma$  parameterized by arc-length has corner points at  $C = \{c_1, \dots, c_k\} \subset [0, 1]$ . Let  $\arg \dot{\gamma}$  denote the angle between the tangent to  $\gamma$  and the positive  $x$ -axis. Then the curvature term for  $\gamma$  is given by

$$\alpha \int_{\gamma \setminus C} \ddot{\gamma}^2 ds + \beta \sum_{c \in C} |\arg \dot{\gamma}(c_+) - \arg \dot{\gamma}(c_-)|.$$

In other words, the penalty for a corner is proportional to the change in tangent direction at the corner.

## Elastica

Minimizing  $E_{2.1}$  involves interpolating hidden edges behind occluding objects. Although this may seem a rather fanciful pursuit, Kanizsa's experiments [2] have revealed that people often make such interpolations automatically and without thinking whether these imagined curves are justified by the visible parts of the scene. For instance, subjects may report "seeing" partly occluded objects contradicting facts one knows about the shapes and sizes of common objects<sup>2</sup>. These human interpolations are hard to predict and probably involve constructions based on a complex set of remembered shapes. Here we only attempt to model the simplest aspects of this. Suppose a region  $R_i$  disappears behind occluding objects at  $P_0 \in \partial R_i$  and reappears at  $P_1$ . Suppose  $\mathbf{t}_0$  and  $\mathbf{t}_1$  are the unit tangent vectors to  $\partial R_i$  at  $P_0$  and  $P_1$ . Then we can characterize the invisible portion  $\Gamma$  of  $\partial R_i$  between  $P_0$  and  $P_1$  as the curve which minimizes:

$$\int_{\Gamma} (\nu + \alpha \kappa^2) ds$$

subject to the boundary conditions of beginning at  $(P_0, \mathbf{t}_0)$  and ending at  $(P_1, \mathbf{t}_1)$ .

This particular variational problem has a long history, having been first investigated by Euler in 1744<sup>3</sup>. The curves which minimize this integral have been called "elastica" since then, and have appeared here and there, esp. in treatises on elasticity (cf. [7], Chapter 19 and [8] for a recent treatment). One reason why they are not better known is that they are not expressible by simple functions, but require elliptic or similar functions.

Putting complex coordinates in the plane, they may, for instance, be written as logarithmic derivatives of the theta function:

$$F(z) = \frac{\partial}{\partial z} \log \theta(\omega, x + a) - b \cdot x$$

where

$$\theta(\omega, x) = \sum_n e^{\pi i n^2 \omega + 2\pi i n x}$$

and either

$$\begin{aligned} \omega = it, \quad a = \frac{it}{2}, \quad b = F'(\frac{1}{2} - \frac{it}{4}), \quad \text{some real } t, \text{ or} \\ \omega = \frac{it+1}{2}, \quad a = 0, \quad b = F'(\frac{1}{4} + \frac{it}{4}), \quad \text{some real } t. \end{aligned}$$

These forms generate the minimizing curves for any  $\nu$  and  $\alpha$  including those that make loops, i.e. which

<sup>2</sup>[2] p. 40-41 and p.88.

<sup>3</sup>*De Curvis Elastica*, an appendix to [6].

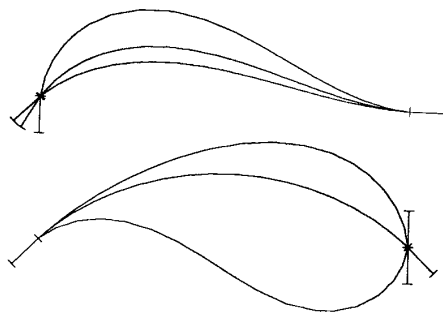


Figure 5: Several elastica computed as described.

are minima in the topological class of curves with each possible total turning angle between  $P_0$  and  $P_1$ . Some examples are shown in figure 5.

Computationally, the simplest way to solve for them seems to be hill-climbing: start with a convenient chain  $P_0 = \mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_N = P_1$  of points joining  $P_0$  and  $P_1$  for which  $\mathbf{x}_1 - \mathbf{x}_0 = \text{constant} \cdot \mathbf{t}_0$  and  $\mathbf{x}_{N-1} - \mathbf{x}_N = \text{constant} \cdot \mathbf{t}_1$  and let it evolve to decrease  $\int (\nu + \alpha \kappa^2) ds$ . Thus we may estimate the curvature  $\kappa_i$  at  $\mathbf{x}_i$  as:

$$\kappa_i = \frac{\theta_i - \theta_{i-1}}{(d_i + d_{i-1})/2}$$

where  $\mathbf{x}_{i+1} - \mathbf{x}_i$  has length  $d_i$  and orientation  $\theta_i$ . Its second derivative can be estimated by:

$$\kappa_i^{**} = \frac{2\kappa_{i+1}}{d_i(d_i + d_{i-1})} - \frac{2\kappa_i}{d_i d_{i-1}} + \frac{2\kappa_{i-1}}{d_{i-1}(d_i + d_{i-1})}$$

Then the curve evolves by:

$$\text{new } \mathbf{x}_i = \mathbf{x}_i + \epsilon(\nu \kappa_i + \alpha \kappa_i^3 - 2\alpha \kappa_i^{**}) \cdot (-\sin(\theta_i), \cos(\theta_i)),$$

for  $2 \leq i \leq N-2$  (n.b.  $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_{N-1}$  and  $\mathbf{x}_N$  stay fixed).

## A Segmentation Algorithm

We outline a segmentation algorithm that finds the 2.1-D sketch of an intensity image  $I$  in three stages: finding edges and T-junctions, hypothesizing continuations, and minimizing  $E_{2.1}$  combinatorically with these edges and continuations.

We have so far developed and coded only the third stage in order to test the concepts behind  $E_{2.1}$  and find whether this functional may be expected to give reasonable results in general. The first two stages are the subject of current investigations.

The first stage finds intensity edges using a common algorithm such as that of Canny [9]. It assembles edge points into curves using estimates of tangent and curvature in a neighborhood larger than nearest neighbors on a grid. We are experimenting with a simple curvature estimate based on the observed behavior of end-stopped cells in the primate visual system. This gives a set of curves  $\Gamma = \{\gamma_1, \dots, \gamma_k\}$  which do not intersect. Next, their endpoints are joined to nearby curves, both to jump gaps and to form T-junctions and corners. Corners and T-junctions can be confirmed using a corneriness measure, such as

$$\left\| \begin{pmatrix} I_{xx} & I_{xy} \\ I_{xy} & I_{yy} \end{pmatrix} \begin{pmatrix} -I_y \\ I_x \end{pmatrix} \right\|^2,$$

described by Kitchen and Rosenfeld in [10]. Along edges, it takes maxima at points with the greatest 1-dimensional curvature. In these cases, the endpoints of the curves,  $\gamma_i(0)$  and  $\gamma_i(1)$ , are tagged as corners and T-junctions for the second stage.

To complete curves across the gaps left by the edge-detector, we put a prior on the space of curves, with fixed tangent and possibly also curvature at the endpoint. The prior can be a simple heuristic function that encourages completions across short, straight gaps, with increasing penalty as curvature increases. Very interesting work in the same direction has been carried out by Zucker *et. al.*, see [11].

For each completion that does not correspond to a T-junction, we delete the two fragments of curves from  $\Gamma$  and add the completed curve to  $\Gamma$ . At T-junctions and corners, we split the curves, so that at the end of the first stage of the algorithm,  $\Gamma = \{\gamma_1, \dots, \gamma_h\}$  is a set of curves which meet each other only at their endpoints.

The second stage hypothesizes new contours from edge termination cues, by matching up pairs of endpoints, disqualifying unlikely matches via simple heuristics, and computing continuations for the remaining candidate pairs.

One heuristic requires the length scale of a continuation to be reasonably close—say, within an order of magnitude—to the lengths of the two curve segments being joined. Thus if two curves of length 1 are 15 units apart, they should be disqualified. We feel this is an important heuristic and should be substantiated with further empirical evidence. Another heuristic requires that matching T-junctions have very similar intensity on corresponding sides of the base of the T. A third heuristic disqualifies a triple point if it contains a cusp; these arise from markings on a curved surface, and should not be continued. Near a triple point, it is more difficult to approximate the curvature of one edge due to its interaction with the others nearby.

Continuations are computed using a prior, which in this case should give an approximation to the elastica described earlier. The new curves  $\{\gamma_{h+1}, \dots, \gamma_g\}$  are added to  $\Gamma$ , and are marked as “hypothetical.”

The third and final stage minimizes  $E_{2.1}$  on  $I$  by combinatoric search, to find the optimal segmentation wherein all region boundaries lie along contours in  $\Gamma$ .

Given  $\Gamma = \{\gamma_1, \dots, \gamma_g\}$ , a set of non-self-intersecting contours which meet each other and  $\partial D$  only at endpoints, we let  $\Pi = \{P_1, \dots, P_n\}$  be the closures of the connected components of  $D \setminus \Gamma$ . The algorithm solves the problem of finding an overlapping segmentation  $(\{R_i\}, <)$  of  $D$  which minimizes  $E_{2.1}$  such that  $\partial R_i \subset \cup \Gamma$ .

In this setting, the algorithm needs only a finite number of scalar values as input, together with an embedded oriented planar graph whose edges are the  $\gamma \in \Gamma$ , which we orient arbitrarily. The values necessary, for each  $\gamma \in \Gamma$  and for each  $P \in \Pi$ , are:

- 1  $|\gamma \setminus \partial D|$ , the length of the contour;
- 2  $\int_{\gamma \setminus \partial D} \phi(\kappa) ds$ , the curvature term of  $E$  computed on the contour except where it coincides with  $\partial D$ ;
- 3  $\arg \dot{\gamma}(0)$  and  $\arg \dot{\gamma}(1)$ , the orientations of the tangents at the endpoints of the contour;
- 4  $\bar{g}_P$ , the *mean* value of an image function  $g : D \rightarrow \mathbf{R}$  on  $P$ ; and
- 5  $|P|$ , the area of  $P$ .

The vertices  $V$  of the embedded graph are the set of points in  $D$  which are endpoints of some  $\gamma$ . To specify the graph structure, it suffices to associate, to each contour  $\gamma$ ,

- 6  $v_{0,\gamma}$  and  $v_{1,\gamma}$ , the vertices it links, and
- 7  $P_{0,\gamma}$  and  $P_{1,\gamma}$ , the faces it separates,

labelled so that as one travels from  $v_{0,\gamma}$  to  $v_{1,\gamma}$ ,  $P_{0,\gamma}$  is on one's left and  $P_{1,\gamma}$  on one's right.

Given these as input, the general strategy is to compute optimal segmentations for subsets of  $\Pi$  whose unions have connected interior, starting with just single  $P$ 's, then in two's, three's, and so on, each time using the previous results. This performs the desired combinatoric search, and gives the 2.1-D sketch for the image, given the curves and continuations from the first two stages.

We have applied the combinatoric search algorithm to the examples in figures 1(a) and 3(a). The contours  $\Gamma$  which we used are those indicated in figures 1(b) and 3(b): in the case of figure 3, we have added several plausible candidates for the extensions of the boundaries of the visible regions. In both cases  $\Pi$  consists of

11 disjoint regions  $P_i$ , while  $\Gamma$  has 25 edges in figure 1(b) and 18 edges in figure 3(b). In both cases, the algorithm gives the intuitively "correct" overlapping segmentation shown in figures 1(c) and 3(c), provided the constants are correctly chosen.  $\epsilon$  is unimportant so long as it is relatively small. The other constants need to satisfy a rather complex set of constraints:

1. The penalty  $\beta$  for angles must be sufficiently large compared to  $\mu$  and  $\nu$  because introducing overlapping regions trades fewer angles for greater intensity variance and contour length.
2. On the other hand, if  $\mu$  is too small compared to  $\nu$  and  $\beta$ , the energy will be minimized by the null segmentation: no regions except for the background.
3. If  $\mu$  is too large compared to  $\nu$  and  $\beta$ , then the energy will be minimized by a segmentation with many many contours, which follow small texture features as well as true object boundaries.

A compromise which seems to work in most cases is this. First compute the total variance of the image, without any segmentation. Call this tot-var. Then to get reasonable results out of the non-overlapping  $E_{M-S}$ , we can set:

$$\nu = \frac{\text{tot-var}}{40\text{diam}(D)}\mu^2.$$

(This may be justified heuristically by assuming that the contours may remove about half the total variance and that to do so we may need about 10 horizontal and 10 vertical lines.) For  $E_{2,1}$ , the curvature coefficients may be set by:

$$\alpha = \frac{\text{tot-var} \cdot \max(\text{curv})}{60}\mu^2$$

$$\beta = \frac{\text{tot-var}}{60}\mu^2$$

where  $\max(\text{curv})$  is the maximum of the expected curvature of the contours which you want to preserve (i.e. above this limit, the quadratic penalty turns to a linear one). The heuristic for  $\beta$  is to roughly equalize the penalty for a right angle and a contour running across the image. These gave us values which worked for figures 1 and 3.

## References

- [1] D. Marr, *Vision*, New York: W H Freeman and Company, 1982.
- [2] G. Kanizsa, *Organization in Vision*, New York: Praeger, 1979, Ch. 1-2.
- [3] R. von der Heydt, E. Peterhans and G. Baumgartner, "Illusory Contours and Cortical Neuron Responses," *Science* **224** 1260-1262, 1984.
- [4] K. Nakayama and S. Shimojo, "Towards a neural understanding of visual surface representation," *Cold Spring Harbor Symposium on Quantitative Biology*, Volume 55, The Brain, Edited by T. Sejnowski, E.R. Kandel, C.F. Stevens and J.D. Watson, to appear 1990.
- [5] D. Mumford, and J. Shah, Optimal Approximations of Piecewise Smooth Functions and Associated Variational Problems, *Comm. in Pure and Appl. Math.*, 1989, **42**, pp.577-685.
- [6] L. Euler, *Methodus inveniendi lineas curvas maximi minimive proprietate gaudentes*, Lausanne, 1744.
- [7] A. E. H. Love, *A treatise on the mathematical theory of elasticity*, Cambridge: The University Press, 1934.
- [8] R. Bryant, and P. Griffiths, Reduction for constrained variational problems and  $\int \frac{\kappa^2}{2} ds$ . *Amer. J. Math.*, **108** (1986), p.525.
- [9] J. F. Canny, *Finding Edges and Lines in Images*, MIT AI Technical Report No. 720, June 1983.
- [10] L. Kitchen and A. Rosenfeld, *Gray-Level Corner Detection*, Technical Report TR-887, Computer Science Center, University of Maryland, College Park, 1980.
- [11] C. David and S. W. Zucker, *Potentials, Valleys, and Dynamic Global Coverings*, McGill University Technical Report TR-CIM 89-1, March 1989.