# Seeds of Perception

Michael Brady
Robotics Research Group,
Department of Engineering Science,
University of Oxford,
Oxford OX1 3PJ,
U.K.

## 1 Abstract

Not all information is created equal. Different locations in an image (or range image) impose differing levels of constraint upon visual processes. We call locations in images (and other representations) that offer the tightest constraint *seeds*. We are exploring algorithms that work initially from seeds to locations of ever decreasing constraint. Thus in motion, we work from corners to edges and then to regions bounded by those edges. We recount recent progress in developing this theme in the problems of computing shape from contour, the determination of optic flow and structure from motion, and stereo. We review related work in shape from shading, and the interpretation of variable geometry planar shapes. The loci of two dimensional change that are discussed in this paper are uncovered by the programs of Fleck and Noble at this conference, as well as by Forsyth's colour labelling of edges.

## 2 Introduction

Not all information is created equal. Different locations in an image (or range image) offer different levels of constraint to visual processes. This observation was first studied, from an information theoretic standpoint, by Attneave (Attneave 55) who noted that line junctions and locations of high curvature on a cartoon drawing of a cat were particularly effective in eliciting fast, accurate recognition. In turn, Attneave's observation inspired our earlier work on the *Curvature Primal Sketch* (Asada and Brady 86) in which we proposed a formal model for the isolation of sharp changes of curvature along image contours. The local constraints afforded by image junctions in images of polyhedra were uncovered in early work on line drawing interpretation (Huffman 71) (Clowes 71) and more recently for line-drawings of images of curved objects (Horaud and Brady 87) (Binford 81) (Malik 87).

Generally, a region of an image that is the projection of a portion of a smooth surface imposes less constraint on visual interpretation than do one-dimensional loci of change, which often correspond to visual edges. In the following, we shall refer to loci of one-dimensional intensity change as "edges" despite the fact that the step edges ubiquitously studied in computer vision are only one of a number of important kinds of change (Canny 83) (Ponce and Brady 87) (Asada and Brady 86). Edges offer even more constraint if they can be elaborated by descriptions that relate to important visible surface characteristics such as texture or colour (Forsyth 87). We call regions that correspond to smooth surface patches *loci of zero-dimensional change*, and they are roughly characterized by two large eigenvalues in the image autocorrelation function as the image "looks" the same in all directions (more on this below). Similarly, we call edges *loci of one-dimensional change*. The autocorrelation function has a large eigenvalue lying along the edge and a small eigenvalue in the direction of the normal to the edge.

By analogy, we refer to corners, points of occlusion (eg T-junctions and X-junctions), and various curvature maxima as *loci of twodimensional change*. Attneave's observation, as well as the work referred to above of Brady and Asada, Clowes, Huffman, Binford, and Malik, amount to a statement that one-dimensional loci of change impose less constraint than twodimensional loci of change. The parameters of a visual process can often be determined not only more completely but also more *reliably* at loci of twodimensional change than at loci of one-dimensional change. In turn, those visual parameters can be computed more completely or reliably at loci of one-dimensional change than at smooth or little change. Often, indeed, there is so little constraint on the values of visual parameters available at smooth portions that the computation has to be supported by regularisation processes, which, to varying extents allow smoothness conditions to dominate the information actually available from the data.

This observation suggests that the computation of the parameters of a visual process might be computed most effectively by basing that computation first on loci of twodimensional change, then on loci of one-dimensional change, and finally on loci of smooth change. Indeed, it seems possible that models can be invoked very early in this computation, possibly even on the basis of the determination of the most constraining information available in the image. This suggests in turn that the refinement of model invokation and the determination of rich image descriptions can proceed hand in hand. According to this broad scheme, loci of twodimensional change are *seeds of perception* from which are grown descriptions of one-dimensional change and, eventually, image descriptions. The purpose of this paper is to explore this theme in a number of visual problems: shape from contour; optic flow and structure from motion; stereo; shape from shading; and model-based recognition of objects.

Unfortunately, the reliable computation even of loci of one-dimensional changes is quite difficult. Nevertheless, despite a continuing stream of objection by some computer vision researchers that reliable edge detection is an unrealisable goal, recent work has shown that the advantages to be gained from intensive study of intensity and range changes (Boie, Cox, and Rehak 86) (Shen and Castan 86) (Ponce and Brady 87) (Canny 86) (Spacek 87) is worth the computational expense, especially as the computation is inherently local and well suited to computation on SIMD processors (Page 87). By the way, we note that existing edge detection designs, including directional edge finders such as Canny's as well as non-directional edge finders such as Marr-Hildreth's Laplacian-of-a-Gaussian, effectively concentrate on one-dimensional changes. The Canny edge detector often fails to locate corners, T-junctions, and X-junctions (see (Noble 87) for a review). Though they do not concentrate on step changes in intensity or range, (Brady, Ponce, Yuille, and Asada 85) and (Ponce and Brady 87) assume that image surfaces and range maps are locally cylindrical, that is that change is restricted to a single direction. Note that their analysis does not extend to loci of twodimensional change. Because they are theoretically and practically restricted to loci of one-dimensional intensity change, edge finders are essentially similar to schemes for the detection of temporal changes in conventional signal processing. Loci of twodimensional change are, however, *unique* to image processing, and inevitably require analysis using differential geometry and differential topology (Fleck 87) (Noble 87).

The reliable computation of loci of twodimensional change (which, by analogy we call "corners", though twodimensional changes assume many forms) has often been judged difficult because of their susceptibility to noise. Certainly, there has been less work on the problem of corner detection than on edge detection, and published schemes leave much to be desired (Noble 87). However, it seems premature to give up or even to be pessimistic. Indeed, we shall present evidence to suggest that loci of twodimensional change can be computed reliably, even efficiently. This will allow us to elaborate upon our general theme of the importance of "corners" as the seeds of early vision.

In this paper, we consider a number of visual processes: shape from contour, the determination of optic flow, and struc-

AVC 1987 doi:10.5244/C.1.36

ture from motion. In each case, we review the vision literature that supports and informs our theme of seeds of perception. Also, in each case we state or prove a result that makes explicit the constraint available at loci of twodimensional change, and we discuss the implications for the design of reliable and efficient algorithms.

## 3 The computation of shape from contour

Analysis of the computation of shape from contour has concentrated mostly on smooth, planar contours (Brady and Yuille 84) (Lee, Seidmann, and Rosenfeld 86) (Witkin 81) (Kanade 81). However, the determination of shape from contour is most effective when there are curvature discontinuities along a curve. Since (Asada and Brady 86) had proposed a method for the detection and localisation of curvature discontinuities along image contours, they applied their method to the determination of shape from contour.

Consider a planar curve $\gamma(s)$ that is imaged after undergoing a general affine transform $\mathbf{T}$, that is, a translation, rotation, and scaling. Since affine transforms, as well as the orthographic or perspective image projection $\mathbf{P}$ are linear operations, the zero-crossings and curvature extrema of $\gamma$ appear as zero-crossings and curvature extrema of the image of $\mathbf{P}(\mathbf{T}(\gamma))$ (see also (Marr 77) (Huttenlocher and Ullman 87)). In general, metric quantities such as angles, lengths, and ratios of lengths associated with the shape $\gamma$ are not preserved in $\mathbf{P}(\mathbf{T}(\gamma))$, and so they are of limited usefulness for determining shape from contour. There are, however, at least two constraints that *can* be used to effect the determination of $\gamma$ from its oriented projection $\mathbf{P}(\mathbf{T}(\gamma))$:

1. *the order constraint*: the order of curvature changes around the projected planar shape is unaffected by affine transformation. Other shapes may occlude the shape of interest, as for example when an aeroplane is partly occluded by cloud. In such cases, curvature changes not associated with the shape of interest may obtrude into the curvature change sequence; nevertheless, subsequences of curvature changes associated with the shape to be recognised are generally visible.

2. *the type constraint*: there are many different types of curvature change; for example, Asada and Brady consider *corner, crank, end, smooth join, bump* and *dent*, whereas (Hoffman and Richards 82) propose a number of different *codon types* for representing curvature changes. Under a broad range of values for the scaling factor of the affine transformation (determined by the object shape and by the viewing conditions), the type of a curvature change of $\gamma$ is the same as that of the corresponding curvature change in $\mathbf{P}(\mathbf{T}(\gamma))$. For example, in Asada and Brady's notation, the projection of a transformed crank change is typically a crank.

Asada and Brady implemented an unpublished program that recognised a variety of shapes (aeroplanes) that had undergone affine transformation and partial occlusion. The model was represented as a sequence of curvature changes, each with an associated type. Recognition consisted of a Waltz-like labelling of curvature changes to match subsequences of the curvature changes found in an image to subsequences of the model. Significantly, though not surprisingly, composite types such as cranks and ends were most effective for constraining the subsequence match.

In related work, (Turney, Mudge, and Volz 85) recognised overlapped lamina objects by matching portions of the boundary of a model to a portion of the boundary of an overlapped object. The match was most effective for what they called *salient features* of the contour of the shape as determined by a least squares computation, effectively an autocorrelation of the contour as described by its $s - \theta$ representation, where $s$ is the (unit speed) contour length. Salient points correspond to significant changes of curvature. In (Turney, Mudge, and Volz 85) the salient features were templates computed from the actual model contour rather than explicit curvature change model instances and so were inherently incapable of recognising scaled instances.

Shape matching and recognition based on local salient features by Asada and Brady and by Turney, Mudge, and Volz may be contrasted with the approach of Grimson and Lozano-Pérez (Grimson and Lozano-Pérez 86). They argue that matching based on salient features is unreasonable when data is noisy (precluding the accurate computation of salient features) or when objects are overlapped (salient features are more likely to be occluded than gross features). There is a law of excluded middle operating here: salient features cannot be relied upon *solely* for recognition, for the reasons advanced by Grimson and Lozano-Pérez. However, recognition can be more effective, reliable, and efficient if such salient features are available. Recently, working in our Laboratory, (Stein 87) implemented a recognition program that was a blend of (Turney, Mudge, and Volz 85) and (Grimson and Lozano-Pérez 86). Figure 1 shows a typical recognition result obtained by Stein's program.

## 4 The computation of visual motion

The computation of visual motion refers to several distinct though related problems, including: optic flow or short range motion; long range motion (Ullman 79); and the computation of structure from motion. In this section, we discuss these problems in turn from the standpoint of *seeds of perception*.

Most work on optic flow is based on the *motion constraint equation*:

$$\mathbf{N} \cdot \mu = \frac{I_t}{\|\nabla I\|} \qquad (1)$$

which is a first-order Taylor's series expansion of $I(\mathbf{x}+\delta\mathbf{x}, t+\delta t)$ (Nagel 87). In this equation, $\mu(\mathbf{x})$ is the optic flow field and $\mathbf{N}$ is the image unit normal $\nabla I/\|\nabla I\|$.

In their paper introducing the motion constraint equation, (Horn and Schunck 81) noted that it only determines the component of the flow field in the direction of the image gradient. Since the component in the orthogonal direction, which we shall denote by $\mathbf{T}$ for reasons that will become obvious, is unspecified,

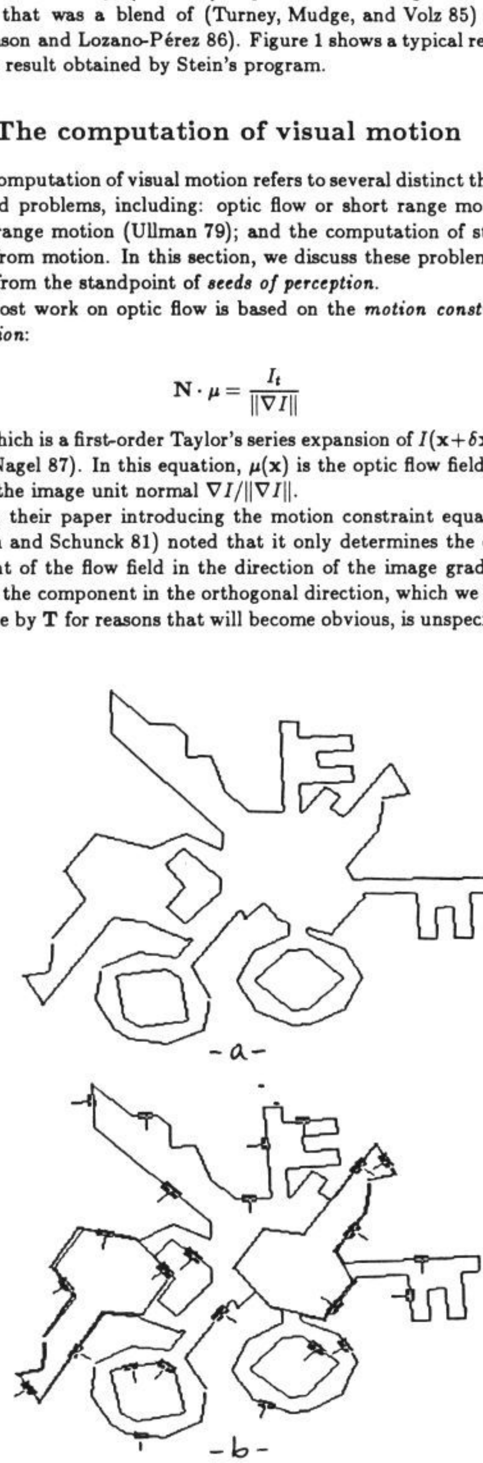


Figure 1: a. Outline of a set of overlapped shapes. b. Key shapes found in the pile shown in a.

the optic flow field is underconstrained. Horn and Schunck suggested adding a smoothness term to regularise the flow. Their algorithm performs well on textured image flows in which there is enough variation in the direction of the image gradient, and hence enough global information to determine the flow field. It performs poorly on motion boundaries.

Since the image gradient $\|\nabla I\|$ occurs in the denominator of the motion constraint equation, the computation of $\mathbf{N}\cdot\mu(\mathbf{x})$ is poorly conditioned unless $\|\nabla I\|(\mathbf{x})$ is large. Often, points $\mathbf{x}$ at which the gradient is large correspond to edges. This is the basis of Hildreth's (Hildreth 84) scheme for the computation of visual motion. Her algorithm identifies points $\mathbf{x}$ at which the image gradient is large with zero-crossings of a Laplacian of a Gaussian filter (Marr and Hildreth 80). The main novelty of her approach is based on a theorem that states that if $\gamma(s)$ is a closed contour (as zero-crossing contours must be unless they are thresholded or cross the image boundary) then there is a unique optic flow field $\mu(\mathbf{x})$ that minimises the smoothness expression:

$$\oint_\gamma \left(\frac{d\mu}{ds}\right)^2 ds$$

so long as the optic flow is different at at least two different points along the closed curve. Hildreth combines the edge smoothing term with the motion constraint equation using a Lagrange multiplier, and solves the resulting minimisation problem using a conjugate gradient descent algorithm. This minimisation algorithm is inherently sequential (Gong 87), rendering it difficult to implement Hildreth's algorithm in parallel. One way to proceed is to replace the conjugate gradient algorithm by a scheme such as Terzopoulous' multigrid relaxation algorithm (Terzopoulos 83) or graduated non-convexity (Blake and Zisserman 87). An alternative approach is described below.

Other authors have noted that even more constraint on the optic flow field is available at loci of twodimensional change. Nagel (Nagel 87) (Dreschler and Nagel 82) has shown that, in practice as well as principle, the *full* optic flow field is computable at what he calls "grey-level corners". Noble (Noble 87) critically analyses Nagel's grey-level corner model. Recently, Nagel (Nagel 86) (Nagel and Enkelmann 86) has shown that the smoothness assumptions underlying the algorithms of Horn and Schunck, of Hildreth, and others amount to special cases of an "oriented smoothness" assumption that is implicit in the use of higher order derivatives. Our work in this section is based on Nagel's. Here we state and prove a generalisation of Nagel's result.

We concentrate on *level contours* of which edges are a particular case. A level contour $\gamma(s)$ is a curve along which the intensity is constant (a contour line on the image surface). As usual, if $\gamma(s)$ is a level contour lying on an image $I(\mathbf{x})$, we denote image points on the level contour by $I(s)$. By the chain rule:

$$\begin{aligned}\nabla I\cdot\frac{d\gamma}{ds} &= \frac{dI}{ds}\\ &= 0\end{aligned} \tag{2}$$

so that along a level contour the image gradient is parallel to the intensity normal $\mathbf{N}$. First, we prove a lemma:

**Lemma**. Along a level contour,

$$\frac{dI_t}{ds}+\|\nabla I\|\mathbf{N}\cdot\frac{d\mathbf{T}}{dt}=0$$

Proof. Since we are restricting attention to a level contour $I(s)$,

$$\frac{\partial}{\partial t}\frac{dI}{ds}=0.$$

However,

$$\frac{\partial}{\partial t}\{\nabla I\cdot\mathbf{T}\}=\nabla I\cdot\frac{d\mathbf{T}}{dt}+\frac{\partial}{\partial t}(\nabla I)\cdot\mathbf{T}$$

Now,

$$\begin{aligned}\frac{\partial}{\partial t}(\nabla I)\cdot\mathbf{T} &= [I_{tx}\ I_{ty}]^\top\cdot\mathbf{T}\\ &= I_{tx}\frac{dx}{ds}+I_{ty}\frac{dy}{ds}\\ &= \frac{dI_t}{ds}\end{aligned} \tag{3}$$

and the result follows. □

**Theorem 1.** If $\mathbf{H}(\mathbf{x})$ is the image Hessian, and $\kappa$ is the (planar) curvature of a level contour whose tangent is $\mathbf{T}$ and whose normal is $\mathbf{N}$, then

$$(\mathbf{N}^\top\mathbf{H}\mathbf{T})\mathbf{N}\cdot\mu=\|\nabla I\|\kappa\mathbf{T}\cdot\mu.$$

Since $\mathbf{T}\cdot\mu$ is required to compute the full flow field, it follows that

1. the full optic flow can be computed at any location along a level contour (including zero crossings) at which the curvature is non-zero.
2. the reliability, or numerical conditioning, of the computation of the full flow increases as the curvature increases. Hence it is most reliable at corners.

Proof. First, note that $\frac{d}{ds}\|\nabla I\|=\mathbf{N}^\top\mathbf{H}\mathbf{T}$. Also, $d\mathbf{N}/ds=-\kappa\mathbf{T}$. We differentiate the motion constraint equation with respect to the unit speed parameter $s$:

$$\begin{aligned}0 &= \frac{d}{ds}(\|\nabla I\|\mathbf{N}\cdot\mu+I_t)\\ &= (\frac{d}{ds}\|\nabla I\|)\mathbf{N}\cdot\mu-\|\nabla I\|\kappa\mathbf{T}\cdot\mu\\ &+ \frac{dI_t}{ds}+\|\nabla I\|\mathbf{N}\cdot\frac{d\mu}{ds}\end{aligned} \tag{4}$$

Since $d\mu/ds=d\mathbf{T}/dt$, the Theorem follows. □

If the curvature $\kappa$ along a level contour is zero, the expression on the right hand side of the Theorem is zero. It is useful to check that the left hand side is also zero! Figure 2 shows a straight portion of contour inclined at an angle $\alpha$ to the $x$-axis. Rotating the coordinate frame to $X-Y$ as shown, we note that $\mathbf{N}=[-\sin\alpha\ \cos\alpha]^\top$ and $\mathbf{T}=[\cos\alpha\ \sin\alpha]^\top$. Clearly, $I_{XX}=I_{XY}=0$, so that the image Hessian is given by:

$$\begin{bmatrix} I_{xx} & I_{xy}\\ I_{xy} & I_{yy}\end{bmatrix}=\begin{bmatrix}\cos\alpha & -\sin\alpha\\ \sin\alpha & \cos\alpha\end{bmatrix}\begin{bmatrix}0 & 0\\ 0 & I_{YY}\end{bmatrix}\begin{bmatrix}\cos\alpha & \sin\alpha\\ -\sin\alpha & \cos\alpha\end{bmatrix} \tag{5}$$

so that $\mathbf{N}^\top\mathbf{H}\mathbf{T}=0$, as expected.

More generally, if the normal $\mathbf{N}$ points in a direction $\alpha$ relative to the $x$-axis,

$$\mathbf{N}^\top\mathbf{H}\mathbf{T}=I_{xy}\cos 2\alpha+\frac{I_{yy}-I_{xx}}{2}\sin 2\alpha \tag{6}$$

Rotating the coordinate frame so that $\alpha=0$, we find that $\mathbf{N}^\top\mathbf{H}\mathbf{T}=I_{xy}$. For rotationally symmetric images, $I_{xy}=0$ and so $\kappa=0$ or $\mathbf{T}\cdot\mu=0$. It follows that Theorem 1 is not applicable to a rotating circle or to a circle whose flow field is a pure expansion. If, for example, $d\kappa/ds\neq 0$ then, in principle at least, the full flow field can be computed locally. Again, a sufficient but not necessary condition for this to be the case is that the image location be a corner.

Nagel (Nagel 87) suggests using the full flow field at corners to determine the optic flow elsewhere. This makes sense since the computation is most well conditioned at those points. Dreschler and Nagel write down the second *spatial* derivative and first *temporal* derivative of the flow field *everywhere* in the image. They note the crucial role played by the image Hessian. This is to be expected from elementary fluid mechanics (Landau and Lifschitz 59). Similarly, in a series of theoretical papers, Koenderinck and van Doorn (Koenderinck and van Doorn 75) (Koenderinck and van Doorn 76) (Koenderinck and van Doorn 81) (Koenderinck, van Doorn, and van de Grind 85) (Koenderinck 86)

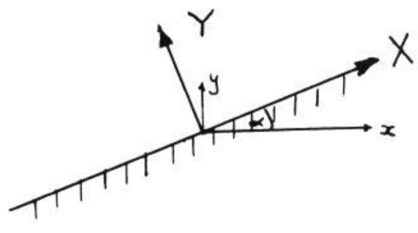


Figure 2: A straight portion of a level contour inclined at angle $\alpha$ to the $x$−axis.

have noted invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. Barnard and Thompson (Barnard and Thompson 82) have pursued a similar approach. Similarly, (Davis, Wu, and Sun 83) have studied the motion of image corners in the relatively easy case of images of moving polyhedra.

Returning to Hildreth's computation, we noted above that zero-crossing contours are closed as is required for the application of her main theorem. In general, zero-crossing contours of a Laplacian-of-a-Gaussian filter do not provide as good an edge map as is produced by a directional operator such as Canny's. In turn, Canny's edge map is not as good, particularly at loci of twodimensional change, as that produced by Fleck's phantom edge finder (Fleck 87). Although Hildreth's theorem relies upon closed contours, it is straightforward to show that:

**Theorem 2**: If the full optic flow is specified as $\mu(s_1) = \mu_1$ and $\mu(s_2) = \mu_2$ at two points $s_1, s_2$ along a level contour, then there is a unique flow field $\mu(s)$ which minimises Hildreth's edge smoothness integral and satisfies the given boundary conditions. □

Theorem 2 can be used to speed the convergence of Hildreth's conjugate gradient scheme by restricting it to portions of contours between loci of twodimensional change at which the full optic flow field can be computed. Between such loci, one might either:

1. Use the Hessian-based formula developed in Theorem 1 directly. This would be a bad idea because the computation of $\mathbf{T} \cdot \mu$ is poorly conditioned when the curvature is small.
2. or, develop a relaxation formula to interpolate the optic flow between known values $\mu_1, \mu_2$ as in Theorem 2.

We are pursuing the latter approach. Note that the situation is trivial for straight edges for which the optic flow linearly interpolates the values at the corners (Murray, Castelow, and Buxton 87). Consider, for example, the following simple scheme:

$$\begin{aligned} \frac{d}{ds}(\mathbf{T} \cdot \mu) &= \frac{d\mathbf{T}}{ds} \cdot \mu + \mathbf{T} \cdot \frac{d\mu}{ds} \\ &= \frac{d\mathbf{T}}{dt} \cdot \mathbf{T} + \kappa\mu \cdot \mathbf{N} \\ &= \kappa\mu \cdot \mathbf{N} \end{aligned} \tag{7}$$

It follows that:

$$(\mathbf{T} \cdot \mu)(s + \delta s) = (\mathbf{T} \cdot \mu)(s) + \delta s\, \kappa (\mathbf{N} \cdot \mu) \tag{8}$$

Note that $\delta s\, \kappa = \delta\alpha$, the incremental change in the tangent angle along the curve.

In a recent conversation, C. Harris noted that this scheme would probably work well for isolated contours (Figure 3a); but that care would be required applying the method to the expected case of contours with junctions indicating overlaps (Figure 3b).

So far, this Section has been concerned with the computation of optic flow. However, work on other problems in the computation of visual motion reinforces our investigation of *seeds of perception*. Ullman (Ullman 79) studied the correspondence computation for isolated *tokens* in long range motion; but he did not state precisely what tokens would correspond to in a real image. Spacek (Spacek 87) proposed a technique for computing image corners using the responses of both a directional and a non-directional edge finder (the responses are related by a simplified version of the Navier-Stokes equation). Corners were matched by weighting the closeness of corners found at time $t+\delta t$ to the expected position, gradient, and curvature of points found at time $t$.

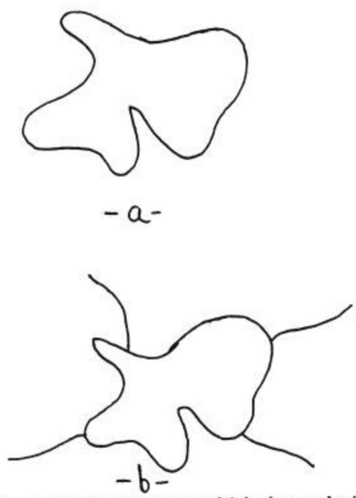


Figure 3: a. An isolated contour to which the analysis developed here may apply straightforwardly. b. A more common case in which junctions arise from overlaps.

Seeds of perception has a place in structure from motion computations too. For example, Ibison, Zapalowski, and Harris (Ibison and Zapalowski 85) determine the three-dimensional layout of scene points by autocorrelation; points "near" corners have higher confidence than points "near" edges, whose confidence is greater than for points within smooth (planar, in fact) surfaces. Harris (this conference) has developed the scheme as follows:

1. Corners are isolated. The technique is analysed by Noble (Noble 87).
2. Corners are matched and error ellipsoids associated with each. The errors are small in the image plane but relatively large in depth.
3. Error ellipsoids are intersected as more views are taken until the estimated position of the feature points found in step 1. are sufficiently good.
4. The estimated three-dimensional positions of the points computed in the previous steps are matched against the vertices of an exact model.

The approach is similar to that of Dreschler and Nagel. We will be working jointly with the Plessey group over the next two years to develop their technique with the work described here and in (Fleck 87) (Noble 87) (Forsyth 87).

## 5 Stereo

In its usual formulation (Buxton 87), the geometry of stereo vision is a simple case of the geometry of 3-d rigid body motion. This assumes that the stereo cameras have previously been rectified, perhaps mechanically. The term "stereo" vision is also coined for stereo-mapping and aerial reconnaisance in which the aeroplane motion between successive images is well approximated by a translation, rotation, and magnification in the image plane. In such applications, rectification may or may not be a separate pre-process prior to stereo matching (Hannah 80) (Gennery 79). Rectification processes typically determine the global image rotation, magnification, and translation by applying a least squares process to a suitable set of sampled points $\mathbf{x}_j^i$, where $i = 1, 2$ denotes the image from which the sample point is taken. The sample set needs to be sparse (for efficiency) and sufficiently constraining to determine the rectification parameters. For precisely the reasons they are used by Harris (this conference), points with low autocorrelation are often chosen as the sample set.

Evidently, rectification can be viewed as a partial stereo match. Indeed, some authors have based stereo algorithms on the sparse point sets that are also suitable for rectification. In perhaps the earliest example of this (Moravec 77) developed an "interest operator" that isolated small image areas with large intensity variation in the four principal directions. His algorithm worked from a coarse to a fine scale choosing the fifty most "interesting" points to match in order to compute a rough range map for his roving vehicle. A variation of Moravec's operator was developed and used by (Hannah 80) in her work on aerial passive navigation and by (Barnard and Thompson 82) in their work on optic flow.

Moravec's stereo matcher restricted attention to a few points for reasons of efficiency. This may not be the only reason to base a stereo matching algorithm upon seeds that are loci of twodimensional intensity change. Recently, Rogers has noted that the (relative) disparity between two matched scene points varies inversely with the square of their depth difference (see (Brady and Hopkins 87) for a precise statement of this). Similarly, the disparity gradient between those points varies inversely with the difference in depth. Rogers has pointed out that the second derivative of disparity (which, by analogy, is called disparity curvature) of those points does not vary with depth, which makes it a particularly attractive parameter on which to base object recognition. In general, the computation of disparity curvature is ill-conditioned. However, as disparity curvature is related to surface curvature, it is best conditioned at loci of two dimensional image change. The computation of such points is also well suited to computation on SIMD processors. We are currently exploring the use of such points as seeds for stereo.

## 6 Related work

Shape from contour and motion are not the only topics that provide support for the notion of *seeds of perception*. In this final section, we briefly mention two other issues: shape from shading; and the interpretation of variable geometry, planar shapes.

The computation of shape from shading of a smooth surface is inherently under-constrained (Horn 86). (Ikeuchi and Horn 81) noted that the surface normal is completely constrained along bounding contours. They used the known values along the bounding contour as seeds in an iterative algorithm that propagated surface normals towards the interior of a shape. (Bruss 81) had earlier explored the uniqueness of the shape from shading problem given a bounding contour. Other work on shape from shading has investigated the substantial constraint offered by highlights, shadow lines, and certain isophotes (Pentland 82), (Blake 85), and (Blake, Zisserman, and Knowles 86).

(Provan 87), working in our Laboratory, has applied truth-maintenance techniques to a problem originally studied in his PhD thesis by (Hinton 76). Hinton's aim was to study the role that relaxation processes might play in constructing structured representations corresponding to perceptual accounts of a scene. As an example, he developed a program to find one or more "optimal" instances of a "puppet" figure in a sea of rectangles. The notion was to deliberately ignore the early processing and concentrate on interpretation. The puppet model is only holonomically constrained: legs are allowed to swing in restricted ways about knees and hips; while arms may swing about elbows and shoulders. In Hinton's approach, the interpretation of a set of overlapped rectangles as an instance of a puppet is initially formulated as an integer programming problem. The integer programming state simply enumerates whether or not the rectangles in the scene belong to a puppet. The state does *not* encode the roles which the selected rectangles play in the corresponding puppet. The symbolic description of how a particular scene can be viewed as containing puppets occurs as a second non-trivial pass.

Provan's program, on the other hand, *explicitly computes the structural description of the puppet at the time that it computes the interpretation*. Provan's program reasons *explicitly* about the puppet structure, thus: I believe rectangle $C$ could be a trunk since rectangle $A$ could be a head, and rectangle $B$ a neck and $A$ is attached to $B$ and $B$ is attached to the putative trunk, and the aspect ratios and position constraints of heads, necks, and trunks are satisfied. The program can recognise puppets in a variety of postures in highly cluttered scenes. If parts of the puppet are missing, for example an upper leg, the program will supply a "default" of the right size and in the right position to join to a putative lower leg. If a more satisfactory upper leg is later added to the image, the interpretation of the lower leg can change automatically.

Despite their considerable differences, both programs start by interpreting a small number of rectangles as putative heads or trunks because they overlap an appropriate number of rectangles that have the right relative aspect ratios to serve as a neck or as upper arms or thighs. These initial interpretations are called "seeds" by Provan and Hinton. The holonomic puppet model then constrains the interpretation of further, intrinsically ambiguous, rectangles as the remaining parts of the puppet.

## 7 Acknowledgements

We thank our several colleagues for many discussions in the development of this paper: Ron Daniel, Hugh Durrant-Whyte, Margaret Fleck, David Forsyth, David Foster, Peter Foulkes, Siao Gan Gong, Radu Horaud, David Murray, Alison Noble, Ian Page, and Guy Scott. This paper would not have been completed without the heroics of Sarah Harrington in guarding the phone.

## References

H. Asada and M. Brady, 1986. The Curvature Primal Sketch. *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-8(1):2–14.

F. Attneave, 1955. Some informational aspects of visual perception. *Psychological Review*, 61.

S. T. Barnard and W. B. Thompson, 1982. Disparity analysis of images. *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-2:333–340.

T. O. Binford, 1981. Inferring surfaces from images. *Artificial Intelligence*, 17:205–245.

A. Blake, 1985. Specular stereo. In *Proc. Int. Jt. Conf. Artif. Intell.*, pages 973–976, Karlsrühe.

A. Blake and A. Zisserman, 1987. *Visual Reconstruction*. MIT Press, Cambridge, Mass.

A. Blake, A. Zisserman, and G. Knowles, 1986. Surface descriptions from stereo and shading. *Image and Vision Computing*, 3:183–191.

R. A. Boie, I. J. Cox, and P. Rehak, 1986. On optimum edge recognition using matched filters. In *Proc. Conf. Pattern Recognition Image Processing*, pages 100–108, Miami Beach, FL.

Michael Brady and P. J. Hopkins, 1987. *Disparity curvature*. Technical Report, Oxford University (in preparation).

Michael Brady and A. Yuille, 1984. An extremum principle for shape from contour. *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-6:288–310.

M. Brady, J. Ponce, A. Yuille, and H. Asada, 1985. Describing surfaces. *Comput. Graphics Image Processing*, 32:1–28.

A. R. Bruss, 1981. *The image irradiance equation: its solution and application*. PhD thesis, MIT Artificial Intelligence Laboratory.

B. F. Buxton, 1987. *Notes on computer vision*. Technical Report, GEC Hirst Research Laboratory.

J.F. Canny, 1983. *Finding Edges and Lines*. Technical Report Tech.Rep. 720, Massachusetts Inst. Technol.

J. Canny, 1986. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-8:679–698.

M. B. Clowes, 1971. On seeing things. *Artificial Intelligence*, 2:79–116.

L. S. Davis, Z. Wu, and H. Sun, 1983. Contour based motion estimation. *Comput. Graphics Image Processing*, 23:313–326.

L. Dreschler and H. -H. Nagel, 1982. Volumetric model and 3-d trajectory of a moving car derived from monocular tv frame sequences of a street scene. *Comput. Graphics Image Processing*, 20:199–228.

Margaret M. Fleck, 1987 (). The phantom edge finder. In *Proceedings of the Alvey Vision Conference*, Cambridge, England.

David Forsyth, 1987. The use of colour in vision. In *Proc. Alvey Vis. Conf.*, Cambridge, UK.

D. B. Gennery, 1979. Stereo camera calibration. In *Proc. Image Understanding Workshop*, page , Science Applications Inc., Silver Springs, Md.

Shaogang Gong, 1987. *Parallel computation of visual motion*. Technical Report , Oxford University MSc report.

W. Eric L. Grimson and Tomás Lozano-Pérez, 1986 (). Search and sensing strategies for recognition and localization of two- and three- dimensional objects. In *Third Symp. Robotics Research*, pages 73–82, Gouvieux, France.

Marsha Jo Hannah, 1980. Bootstrap stereo. In *Proc. Image Understanding Workshop*, page , Science Applications Inc., Silver Springs, Md.

E. C. Hildreth, 1984. *The Measurement of Visual Motion*. MIT Press, Cambridge, Ma.

G. H. Hinton, 1976. *Relaxation and its role in computer vision*. PhD thesis, Edinburgh University.

D. D. Hoffman and W. A. Richards, 1982. Representing smooth plane curves for recognition: implications for figure-ground reversal. In *Proc. Nat. Conf. Artif. Intell.*, pages 5–8, Pittsburgh, Pa.

Radu Horaud and Michael Brady, 1987 (June). On the geometric interpretation of image contours. In *Proceedings of the First International Conf. Comp. Vision.*, pages 374–382, London, England.

B. K. P. Horn, 1986. *Robot Vision*. MIT Press, Cambridge, Ma.

B. K. P. Horn and B. G. Schunck, 1981. Determining optical flow. *Artificial Intelligence*, 17:185–203.

D. A. Huffman, 1971. Impossible objects as nonsense sentences. *Machine Intelligence 6*, 295–323.

D. P. Huttenlocher and S. Ullman, 1987 (June). Object recognition using alignment. In *Proceedings of the First International Conf. Comp. Vision.*, pages 102–111, London, England.

M.C. Ibison and L. Zapalowski, 1985. Structure from motion: an alternative approach. In *Proc. Conf. Pattern Recognition Image Processing*, page 203, .

K. Ikeuchi and B. K. P. Horn, 1981. Numerical shape from shading and occluding boundaries. *Artificial Intelligence*, 17:141–185.

T. Kanade, 1981. Recovery of the three-dimensional shape of an object from a single view. *Artificial Intelligence*, 17:409–461.

Jan J. Koenderinck, 1986. Optic flow. *Vision Research*, 26:161–180.

Jan J. Koenderinck and A. J. van Doorn, 1975. Invariant properties of the motion parallax field due to th emovement of rigid bodies relative to an observer. *Optica Acta*, 22:773–791.

Jan J. Koenderinck and A. J. van Doorn, 1976. Local structure of movement parallax of the plane. *J. Opt. Soc. Am.*, 66:717–723.

Jan J. Koenderinck and A. J. van Doorn, 1981. Exterospecific component of the motion parallax field. *J. Opt. Soc. Am.*, 71:953–957.

Jan J. Koenderinck, A. J. van Doorn, and Wim A. van de Grind, 1985. Spatial and temporal parameters of motion detection in the peripheral visual field. *J. Opt. Soc. Am.*, 2:252–259.

L. D. Landau and E. M. Lifschitz, 1959. *Fluid Mechanics*. Pergamon Press, Oxford.

Chia-Hoang Lee, Thomas Seidmann, and Azriel Rosenfeld, 1986. *A proposed generation process for the reconstruction of space curves*. Technical Report CAR-TR-118, University of Maryland.

J. Malik, 1987. Interpreting line drawings of curved objects. *International Journal of Computer Vision*, 1:73–103.

D. Marr, 1977. Analysis of occluding contour. *Proc. R. Soc. London*, B 197:441–475.

D. Marr and E. C. Hildreth, 1980. Theory of edge detection. *Proc. Roy. Soc. London*, B207:187–217.

H. P. Moravec, 1977. Towards automatic visual obstacle avoidance. In *Proc. Int. Jt. Conf. Artif. Intell.*, page 584, Cambridge, Ma.

D. W. Murray, D. A. Castelow, and B. F. Buxton, 1987. From an image sequence to a recognised polyhedral object. In *Proc. Alvey Vis. Conf.*, Cambridge, UK.

H. -H. Nagel, 1986. Image sequences– ten (octal) years from phenomenology towards a theoretical foundation. In *Proc. 8th Int. Conf. Patt. Recog.*, page , Paris.

H. -H. Nagel, 1987. On the estimation of optical flow. *Artificial Intelligence*, (to appear):.

H. -H. Nagel and W. Enkelmann, 1986. An investigation of smoothness constraints for the estimation of displacement vector fields from image segments. *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-8:565–593.

J.A. Noble, 1987. The Geometric Structure of Images. M.Sc. report.

Ian Page, 1987. *Parallel Algorithms and Computer Vision*. Oxford University Press, Oxford.

A. P. Pentland, 1982. *The visual inference of shape: computation from local features*. PhD thesis, MIT.

J. Ponce and M. Brady, 1987. Towards a surface primal sketch. In *Three-dimensional vision (Kanade ed.)*.

G. M. Provan, 1987. *The complexity of Truth Maintenance Sytems and their application to vision*. PhD thesis, Oxford University.

Jun Shen and Serge Castan, 1986. An optimal linear operator for edge detection. In *Proc. CVPR*, pages 109–114, Miami, Fla.

Libor A. Spacek, 1987. Edge detection and motion detection. *Image and Vision Computing*, 4:43–56.

Fridtjof Stein, 1987. *Recognition of overlapped objects*. Technical Report OU-RRG-87-11, Oxford University Robotics Research Group.

D. Terzopoulos, 1983. The role of constraints and discontinuities in visible-surface reconstruction. In *Proc. 7th Int. J. Conf. Artif. Intell.*, pages 1073–1077, Karlsrühe.

J. L. Turney, T. N. Mudge, and R. A. Volz, 1985. Recognizing partially occluded parts. *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-7:410–421.

Shimon Ullman, 1979. *The Interpretation of Visual Motion.* MIT Press, Cambridge, Ma.

A. Witkin, 1981. Scale space filtering. In *Proc. 7th Int. Jt. Conf. Artif. Intell.*, Karlsrühe, FRG.