Draft dated June 3, 1993

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 1318 December, 1992

Multi-Scale Vector-Ridge-Detection for Perceptual Organization Without Edges

J. Brian Subirana-Vilanova and Kah Kay Sung
Email: subirana@ai.mit.edu, sung@ai.mit.edu

Abstract:

We present a novel ridge detector that finds ridges on vector fields. It is designed to automatically find the right scale of a ridge even in the presence of noise, multiple steps and narrow valleys. One of the key features of such ridge detector is that it has a zero response at discontinuities. The ridge detector can be applied both to scalar and vector quantities such as color.

We also present a parallel perceptual organization scheme based on such ridge detector that works without edges; in addition to perceptual groups, the scheme computes potential focus of attention points at which to direct future processing.

The relation to human perception and several theoretical findings supporting the scheme are presented. We also show results of a Connection Machine implementation of the scheme for perceptual organization (without edges) using color.

Acknowledgements: This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the Laboratory's Artificial Intelligence Research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Army contract DACA76–85–C–0010, and in part by ONR contracts N00014–85–K–0124 and N00014-91-J-4038.

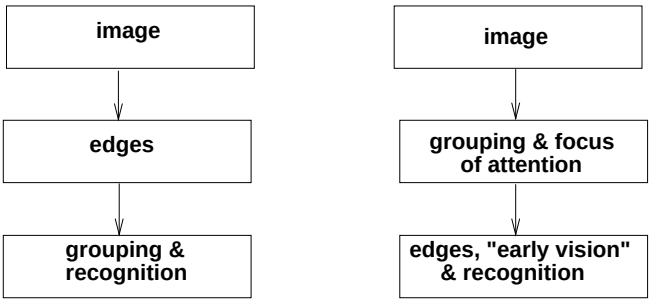


Figure 1: Two different views on the role of perceptual organization.

1 Introduction

Perceptual organization (aka grouping and segmentation) is a process that computes regions of the image that come from different objects, with little detailed knowledge of the particular objects present in the image. Recent work in computer vision has emphasized the role of edge detection and discontinuities in segmentation and recognition. This line of research stresses that edge detection should be done at an early stage on a brightness representation of the image, and segmentation and other early vision modules operate later on (see Figure 1 left). We (like some others) argue against such an approach and present a scheme that segments an image without finding brightness, texture, or color edges (see Figure 1 right). In our scheme, *discontinuities* and a potential *focus of attention* for subsequent processing are found as a byproduct of the perceptual organization process which is based on a novel ridge detector.

Segmentation without edges is not new. Previous approaches fall into two classes. Algorithms in the first class are based on coloring or region growing [Hanson and Riseman 1978], [Horowitz and Pavlidis 1974], [Haralick and Shapiro 1985], [Clemens 1991]. These schemes proceed by laying a few "seeds" in the image and then "grow" these until a complete region is found. The growing is done using a local threshold function, i.e. decisions are made based on local neighborhoods. This results in schemes limited in two ways: first, the growing function does not incorporate global factors, resulting in fragmented regions (see Figure 2). Second, there is no way to incorporate a priori knowledge of the shapes that we are looking for. Indeed, important Gestalt principles such as symmetry, convexity and proximity (extensively used by current grouping algorithms) have not been incorporated in coloring algorithms. These principles are useful heuristics to aid grouping processes and are often sufficient to disambiguate certain situations. In this paper we present a non-local perceptual organization scheme that uses no edges and which embodies these gestalt principles. It is for this reason that our scheme overcomes some of the problems with region growing schemes, mainly the fragmenting of regions and the merging of overlapping regions with similar region properties.

The second class of segmentation schemes which work without edges are based on computations that find discontinuities while preserving some region properties such as smoothness or other physical approximations [Geman and Geman 1984], [Terzopoulos 86], [Blake and Zisserman 1987], [Hurlbert and Poggio 1988], [Poggio, Gamble and Little 1988]. These schemes are scale dependent and in some instances depend on reliable edge detection. Scale has been addressed previously at the discontinuity level [Witkin 1983], [Koenderink 1984], [Perona and Malik 1990] but these schemes do not explicitly represent regions and often meaningful regions are not fully enclosed by the obtained discontinuities. Like with the previous class, all these algorithms do not embody any of the Gestalt principles and in addition perform poorly when there is a nonzero gradient inside a region. The scheme presented in this paper performs perceptual organization (see above) and addresses scale by computing the largest scale at which a structure (*not* necessarily a discontinuity) can be found in the image.

The scheme that we will present is an extension of the brightness-based perceptual organization scheme presented in [Subirana-Vilanova 1990]. Such a scheme is based on a *filter-based ridge detector* with a number of important problems we will discuss. These include its dependence on scale and its sensitivity to curved shapes. Our analysis will lead us to a non-linear filter that overcomes most of these problems.

Our scheme is designed to work for brightness, texture, and color but our implementation deals only with color. Color is an interesting case to study because it is a three-dimensional property, not one-dimensional like intensity making the extension of brightness based schemes to color non-trivial.

We begin in the next section by listing reasons for exploring non-edge based schemes which should give an idea of the difficulties associated with perceptual organization without edges. We then present our approach, including an extended analysis of the ridge-detector, and results of a version of our scheme implemented on the Connection Machine.

2 In Favor of Regions

What is an edge? Unfortunately there is no agreed definition of it. An edge can be defined in several related ways: as a discontinuity in a certain property[1], as "something" that looks like a step edge (e.g. [Canny 1986] - see Figure 3) and by an algorithm (e.g. zero-crossings [Marr and Hildreth 1980]). Characterizing edges has proven to be difficult especially near corners, junctions[2], [Beymer 1991], [Giraudon and Deriche 1991], [Korn 1988], [Noble 1988], [Gennert 1986], [Singh and Shneier 1990], [Medioni and Yasumoto

[1] Note that, strictly speaking, there are no discontinuities in a properly sampled image (or they are present at every pixel)

[2] Junctions are critical for most edge-labeling schemes which do not tolerate well missing junctions.

Figure 2: *(From top-left to bottom right) 1:* Full shirt image. *2:* Canny edges. *3:* Color edges. *4:* An image of a shirt. *5:* Original seeds for a region growing segmentation algorithm. *6:* Final segmentation obtained using a region growing algorithm.

Figure 3: *Left:* Model of an edge. *Right:* Model of a ridge or box. Are these appropriate?

Figure 4: *Left:* Zero-crossings. *Right:* Sign bit. Which one of these is harder to recognize? (Taken from [Marr and Hildreth 1980]).

1987], [Harris and Stephens 1988] and when the image contains edges at multiple scales, noise, transparent surfaces, or edges different than step edges (e.g. roof edges) [Horn 1977], [Ponce and Brady 1985], [Forsyth and Zisserman 1989], [Perona and Malik 1990].

What is a region? Attempting to define regions bears problems similar to those encountered in the definition of an edge. Roughly speaking, it is a collection of pixels in an image sharing a common property. In this context, an edge is the border of a region. How can we find regions in images? We could proceed in a similar way as with edges, so that a region be defined (in one dimension) as a structure that looks like a box (see Figure 3). However, this suffers from problems similar to the ones mentioned for edges.

Thus, regions and edges are two closely related concepts. It is unclear how we should represent the information contained in an image. As regions? As edges? Most people would agree that a central problem in visual perception is finding the objects or structures of interest in an image. These can be defined sometimes by their boundaries, i.e. by identifying the relevant edges in an edge-based representation. However, consider now a situation in which you have a transparent surface as when hair occludes a face, when the windshield in your car is dirty or when you are looking for an animal inside the forest. An edge-based representation does not deal with this case well, because the region of interest is not well defined by the discontinuities in the scene but by the perceived discontinuities. This reflects an object-based view of the world. Instead, a region-based representation is adequate to represent the data in the image. Furthermore, independently of how we choose to represent our data, which structures should we recover first? Edges or regions?

Here are some reasons why exploring the computation of regions (without edges) may be a promising approach:

2.1 *Human Perception*

There is some psychological evidence that humans can recognize images with region information better than line drawings [Cavanaugh 1991]. However, there is not a clear consensus [Ryan and Schwartz 1956], [Biederman and Ju 1988]). See also Figure 4.

2.2 *Perceptual Organization*

Recent progress in rigid-object recognition has lead to schemes that perform remarkably better than humans for limited libraries of models. The computational complexity of these schemes depends critically on the number of "features" used for matching. Therefore, the choice of features is an important issue. A simple feature that has been used is a point of an edge. This has the problem that typically, there are many such features and they

are not very distinctive increasing the complexity of the search process. Complexity can be reduced by grouping this features into lines [Grimson 1990]. Lines in this context are a form of grouping. This idea has been pushed further and several schemes exist that try to group edge segments that come from the same object [Lowe 1984, 1987], [Jacobs 1989]. The general idea underling grouping is that "group features" are more distinctive and occur less frequently than individual features (see [Marroquin 1976], [Witkin and Tenenbaum 1983], [Mahoney 1985], [Lowe 1984, 1987], [Sha'ashua and Ullman 1988], [Jacobs 1989], [Grimson 1990], [Subirana-Vilanova 1990]). This has the effect of simplifying the complexity of the search space. However, even in this domain where existing perceptual organization has found use, complexity still limits the realistic number of models that can be handled. "Additional" groups obtained with region-based computations should be helpful.

Representations which maintain some region information such as the sign-bit of the zero-crossings (instead of just the zero-crossings themselves) can be used for perceptual organization. One property that is easy to recover locally in the sign-bit image shown in Figure 4 is that of membership in the foreground (or background) of a certain portion of the image since a very simple rule can be used: The foreground is black and the background white. (This rule cannot be applied in general, however it illustrates how the coloring provided by the sign bit image can be used to obtain region information.) In the edge image, this information is available but cannot be computed locally. The region-based scheme presented in this paper uses, to a certain extent, a similar principle to the one we have just discussed. Namely, that often regions of interest have uniform brightness properties.

2.3 *Non-rigid objects*

Previous research on recognition has focused on rigid objects. In such a domain, one of the most useful constraints is that the change in appearance, in the image, can be attributable mainly to a change in viewing position and luminance geometry[3]. It has been shown that this implies that the correspondence of a few features constrains the viewpoint (so that pose can be easily verified). Therefore, for rigid-objects, edge-based segmentation schemes which look for small groups of features that come from one object are sufficient. Since cameras introduce noise and edge-detectors fail to find some edges, the emphasis has been on making these schemes as robust as possible under spurious data and occlusion.

Instead, very little research has been devoted to flexible objects such as an alligator. In this case, the change in appearance cannot be attributable solely to a change in viewing direction. Internal changes of the shape have to be taken into account. Therefore, grouping a small subset of image features is not sufficient to recover the object's pose. A different form of grouping that can group all (or most of) the objects features is necessary. Even

[3] For polygonal shapes, in most cases luminance could be ignored if we could recover edges with no errors.

after extensive research on perceptual organization, there are no edge-based schemes that work in this domain (see also the next subsection). This may not be just a limitation on our understanding of the problem but a constraint imposed by the input used by such schemes. The use of more information, not just the edges, may simplify the problem. One of the goals of our research is to develop a scheme that can group features of a flexible object under a variety of settings that is robust under changes in illumination. Occlusion and spurious data should also be considered, but they are not the main driver of our research.

2.4 *Stability and Scale*

In most images, interesting structures in different regions of the image occur at different scales. This is a problem for edge-based grouping because edge detectors are very sensitive to the "scale" at which they are applied. This presents grouping schemes two problems: it is not clear what is the scale at which to apply edge detectors and, in some images, not all edges of an object appear accurately at *one* single scale. Scale stability is in fact one of the most important sources of noise and spurious data mentioned above.

Consider for example Figure 5 where we have presented the edges of a person at different scales. Note that there is no single scale where the silhouette of the person is not broken. For the purposes of recognition, the interesting edges are obviously the ones corresponding to the object of interest. Determining the scale at which these appear is not a trivial task.

This problem has been addressed in the past [Zhong and Mallat 1990], [Lu and Jain 1989], [Clark 1988], [Geiger and Poggio 1987], [Schunck 1987], [Perona and Malik 1987], [Zhuang, Huang and Chen 1986], [Canny 1985], [Witkin 1984] but edge detection has treated scale as an isolated issue, independent of the other edges that may be involved in the object of interest. We believe that the stability and scale of the edges should depend on the region that they belong to and not solely on the discontinuity that gives rise to them. The scheme that we will present looks for the objects directly, not just for the individual edges. This means that in our research we address stability in terms of objects (not edges). In fact, our scheme commits to one scale which varies through the image; usually it varies also within the object. This scale corresponds to that of the object of interest chosen by our scheme.

3 Color, Brightness Or Texture?

The perceptual organization scheme presented in this paper includes color, brightness and texture. We decided to implement it on color first, without texture or brightness. Color based perceptual organization (without the use of other cues) is indeed possible for humans since two adjacent untextured surfaces viewed under iso-luminant conditions can be segmented. (Although the human visual system has certain limitations in iso-luminant

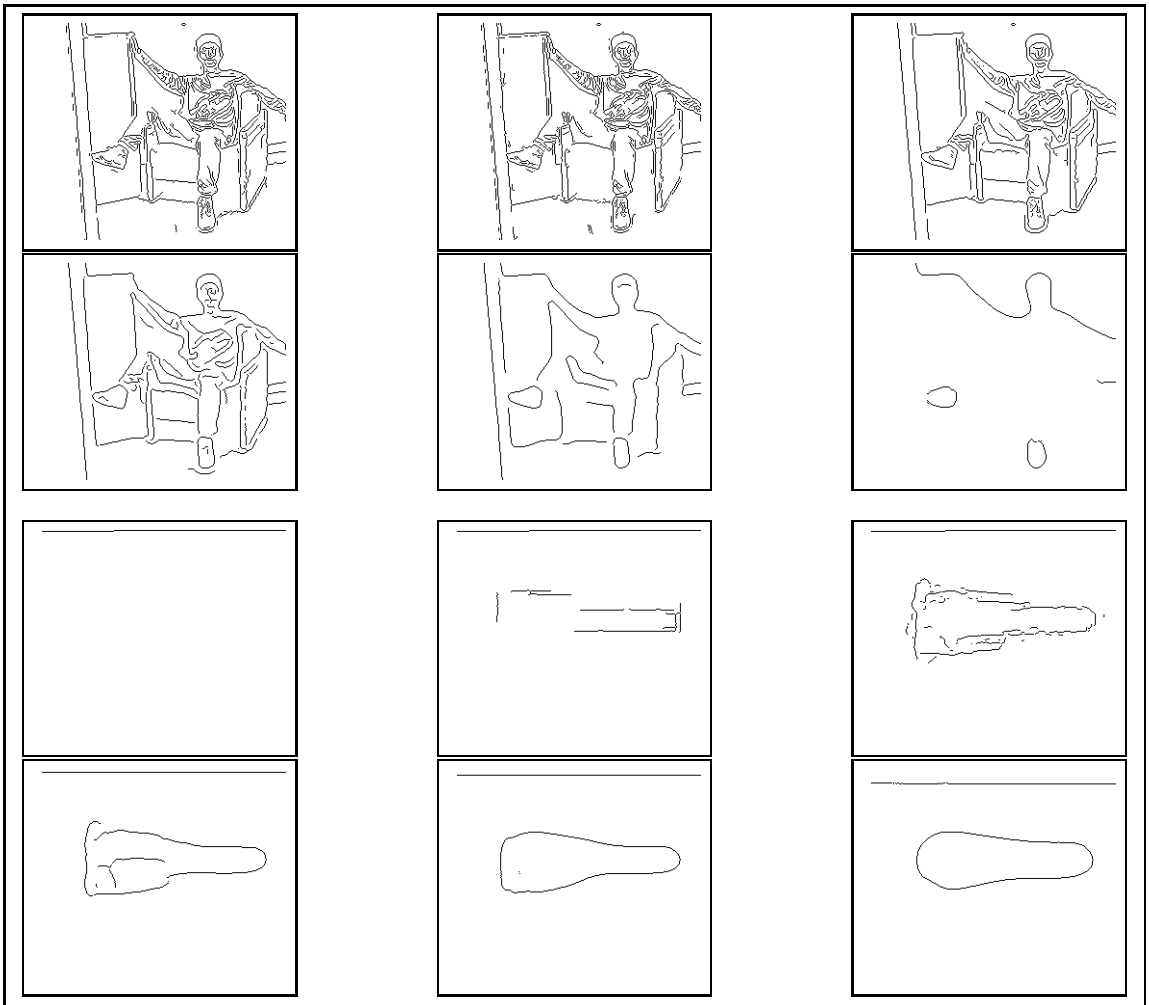


Figure 5: Edges computed at six different scales. Note that the results are notably different. Which scale is best? *Top six:* Image of a person. Note that some of the edges corresponding to the legs are never found. *Bottom six:* Blob image.

displays, e.g. [Cavanaugh 1987].) And, as we will discuss later in the paper, color is also useful when there are brightness changes.

Under normal conditions, color is a perceived property of a surface that depends mostly upon surface spectral reflectance and very little on the spectral characteristics of the light entering our eyes. It is therefore useful for describing the material composition of a surface (independently of its shape and imaging geometry) [Rubin and Richards 1981]. Lambertian color is indeed uniform over most untextured physical surfaces, and is stable in shadows, and under changes in the surface orientation or the imaging geometry. In general it is more stable than texture or brightness. It has long been known that the perceived color (or

intensity) at any given image point depends on the light reflected from the various parts of the image, and not only on the light at that point. This is known as the simultaneous-contrast phenomena and has been known at least since E. Mach reported it at the beginning of the century. [Marr 1982] suggests that such a strategy may be used because one way of achieving some compensation for illuminance changes is by looking at differences rather than absolute values. According to this view, a surface is yellow because it reflects more "yellow" light than a blue surface, and not because of the absolute amount of yellow light reflected (of which the blue surface may reflect an arbitrary amount depending on the incident light).

The exact algorithm by which humans compute perceived color is still unclear. Our scheme only requires a rough estimate of color which is used to segment the image, see Figure 6. We believe that perceived color should be computed at a later stage by a process similar to the ones described in [Helson 1938], [Judd 1940], [Land and McCann 1971]. This model is in line with the ones presented in [Subirana-Vilanova and Richards 1991] and [Jepson and Richards 1991] which suggest that perceptual organization is a very early process which precedes most early visual processing. In our images, color is entered in the computer as a "color vector" with three components: the red, green, and blue channels of the video signal. Our scheme works on color differences $\mathcal{S}_{\otimes}$ between pairs of pixels $\boldsymbol{c}$ and $\boldsymbol{c}_R$. The difference that we used is defined in equation 1 and was taken from [Sung 1991] ($\otimes$ denotes the *vector cross product* operation) and responds very sensitively to color differences between similar colors.

$$\mathcal{S}_{\otimes}(\boldsymbol{c}) = 1 - \frac{|\boldsymbol{c} \otimes \boldsymbol{c}_R|}{|\boldsymbol{c}||\boldsymbol{c}_R|} \tag{1}$$

This similarity measure is a decreasing function with respect to the angular color difference. It assigns a maximum value of 1 to colors that are identical to the reference "ridge color", $\boldsymbol{c}_R$, and a minimum value of 0 to colors that are orthogonal to $\boldsymbol{c}_R$ in the RGB vector space. The discriminability of this measure can be seen intuitively by looking at the normalized image in Figure 6. The exact nature of this measure is not critical to our algorithm. What is important is that when two *adjacent* objects have different perceived color (in the same background) this measure is positive[4]. Many other measures have been proposed in the literature and they could be incorporated in our scheme.

What most color similarity measures have in common is that they are based on vector values and cannot be mapped onto a one-dimensional field [Judd and Wyszecki 75][5]. This makes color perception different from brightness from a computational point of view since

[4]Note that the perceived color similarity among arbitrary objects in the scene will obviously not correspond to this measure. Specially if we do not take into account the simultaneous-contrast phenomena

[5]Note that using the three channels, red, green and blue independently works for some cases. However it is possible to construct cases in which it does not as when an object has two discontinuities, one in the red channel only and the other in one of the other two channels only. In addition, the perceived similarity is not well captured by the information contained in the individual chapels alone but on the combined measure.

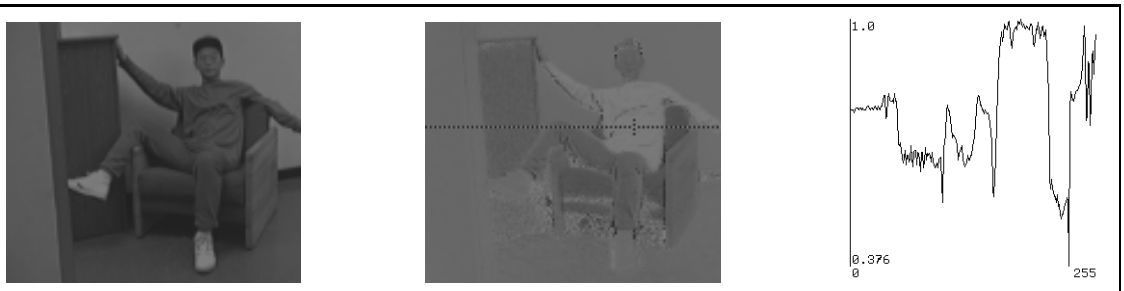


Figure 6: The similarity measure described in Equation 1 is illustrated here for an image of a person. *Left:* Image. *Center:* Similarity measure, using as reference color, the color of the pixel located at the intersection of the two segments shown. *Right:* Plot of the similarity measure along the long segment using the same reference color.

not all the one-dimensional techniques used in brightness images extend naturally to higher dimensions.

4 Regions? What Regions?

In the last two sections we have set forth an ambitious goal: Develop a perceptual organization scheme that works on the image itself, without edges and using color, brightness, and texture information.

But what constitutes a good region? What "class" of regions ought to be found? Our work is based on the observation that many objects in nature (or their parts) have a common color or texture, and are long, wide, symmetric, and convex. This hypothesis is hard to verify formally, but it is at least true for a collection of common objects [Snodgrass and Vanderwart 1980] used in psychophysics. And as we will show, it can be used in our scheme yielding seemingly useful results. In addition, humans seem to organize the visual array using this type of principles as demonstrated by the Gestalt Psychologists [Wertheimer 1923], [Koffka 1935], [Köhler 1940]. In fact, these were the starting point for much of the work in computer vision on perceptual organization for rigid objects. We use these same principles but in a different way: Without edges and with non-rigid shapes in mind.

In the next section we describe some common problems in finding regions. To do so, we introduce a one dimensional version of "regions" and discuss the problems involved in this simplified version of the task. A scheme to solve the one dimensional version of the problem is discussed in Sections 6 and 7. This exercise is useful because both the problems and the solution encountered generalize to the two dimensional version, which is presented in Sections 8 and 9.

5 Problems in Finding Brightness Ridges

One way of simplifying the perceptual organization task is to start by looking at a one dimensional version of the problem. This is especially true if such a solution lends itself to a generalized scheme for the two dimensional problem. This would be a similar path to the one followed by most edge detection research. In the case of edge detection, the generally accepted one dimensional version of the problem is a step function (as shown in Figure 3). Similarly, perceptual organization without edges can be cast in one dimension as the problem of finding ridges similar to a hat (as shown in Figure 3). A hat is a good model because it has one of the basic properties of a region: it is uniform and has a discontinuity in its border. As we will see shortly, the hat model needs to be modified before it can reflect all the properties of regions that interest us.

In other words, the one-dimensional version of the problem that we are trying to solve is to locate *ridges* in a one-dimensional signal. By ridge we mean something that "looks like" a pair of step edges (see Figure 3). A simple-minded approach is to find the edges in the image, and then look for the center of the two edges. This was the approach used in [Subirana-Vilanova 1990]. Another possibility is to design a filter to detect such a structure as in [Canny 1985], [Noble 1988]. This also was the essence of the brightness based approach used in [Subirana-Vilanova 1990].

However, there are a number of problems with using such filters as estimators for ridge detection. These problems are not particular to either scheme, but are linked to the nature of ridges in real images. Some of these problems are in fact very similar for color and for brightness images. The model of a ridge used in these schemes is similar to the one shown in Figure 3. This is a limited model since ridges in images are not well suited to it. Perhaps the most evident reason why such a model is not realistic is the fact that it is tuned to a particular scale, while, in most images, ridges appear at multiple and unpredictable scales. This is not so much of a problem in edge-detection as we have discussed in the previous sections, because the edges of a wide range of images can be assumed to have "a very similar scale". Thus, Canny's ridge detector works only on images where all ridges are of the same scale as is true in the text images shown in [Canny 1983] (see also Figures 17 and 18) and in the images used by [Subirana-Vilanova 1990].

Therefore, an important feature of a ridge detector is its scale invariance. We now summarize a number of important features that a ridge operator should have (see Figure 7):

- *Scale:* See previous paragraph.
- *Non-edgeness:* The filter should give no response for a step edge. This property is violated by [Canny 1985].

Figure 7: *Left:* Plot with multiple steps. A ridge detector should detect three ridges. *Right:* Plot with narrow valleys. A ridge detector should be able to detect the different lobes independently of the size of the neighboring lobes.

- *Multiple steps:* The filter should also detect regions between small steps. These are frequent in images, for example when an object is occluding the space between two other objects. This complicates matters in color images because the surfaces are defined by vectors not just scalar values.
- *Narrow valleys:* The operator should also work in the presence of multiple ridges even when they are separated by small valleys.
- *Noise:* As with any operator that is to work in real images, tolerance to noise is a critical factor.
- *Localization:* The ridge-detector output should be higher in the middle of the ridge than on the sides.
- *Strength:* The strength of the response should be somehow correlated with the strength of the perception of the ridge by humans.
- *Large scales:* Large scales should receive higher response. This is a property used by [Subirana-Vilanova 1990]'s scheme and is important because it embodies the preference for large objects (see also section 14).

6 A Color Ridge Detector

In the previous section we have outlined a number of properties we would like our ridge-detector to have. As we have mentioned, the Canny ridge-detector fails because, among other things, it cannot handle multiple scales. A naive way of solving the scale problem would be to apply the Canny ridge detector at multiple scales and define the output of the filter at each point as the response at the scale which yields a maximum value. This filter would work in a number of occasions but has the problem of giving a response for step edges (since the ridge-detector at any single scale responds to edges, so will the combined filter - see Figures 17 and 18).

One can suppress the response to edges by splitting Canny's ridge operator into two pieces, one for each edge, and then combining the two responses by looking at the *minimum*

Figure 8: *Left:* Gaussian second derivative, an approximation to Canny's optimal ridge detector. *Right:* Individual one-dimensional masks used by our operator.

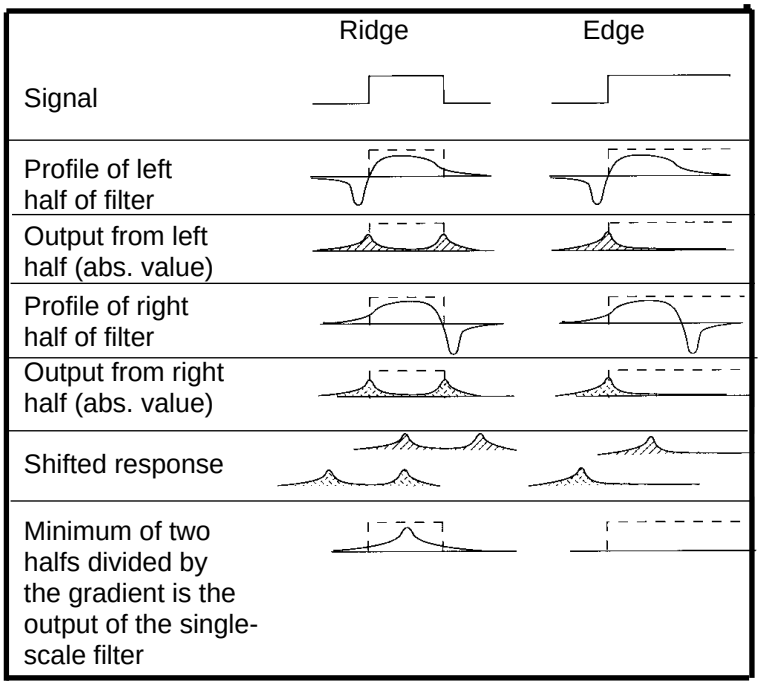


Figure 9: Intuitive description of ridge detector output on flat ridge and edge.

of the two responses. This is the basic idea behind our approach (see Figures 8 and 9). Figures 17 and 18 illustrate how our filter behaves according to the different criteria outlined before. The Figure also compares our filter with that of the second derivative of a gaussian, which is a close approximation to the ridge-filter Canny used. There are a number of potential candidates within this framework such as splitting a Canny filter by half, using two edge detectors and many others. We tried a number of possibilities on the Connection Machine using a real and a synthetic image with varying degrees of noise. Table 6 describes the filter which gives a response most similar to the inertia values and the tolerated length that one would obtain using similar formulas for the corresponding edges, as described in [Subirana-Vilanova 1990].

VAR.	EXPRESSION	DESCRIPTION
$\mathcal{P}_{max}$	Free Parameter (3)	Gradient penalization coeff.
F_s	Free Parameter (8)	Filter Side Lobe size coeff.
F_c	Free Parameter (1/8)	Local Neighborhood size coeff.
$g(x)$		Color gradient at location x.
g_{max}		Max. color gradient in image.
σ		Size of Main Filter Lobe.
σ_s	σ / F_s	Size of Side Filter Lobe.
σ_c	$F_c \sigma$	Reference Color Neighborhood
$\boldsymbol{c}(x)$	$[\mathcal{R}(x)\ \mathcal{G}(x)\ \mathcal{B}(x)]^T$	Color vector at location x.
$\boldsymbol{c}_n(x)$	$\boldsymbol{c}(x)/\lvert\boldsymbol{c}_(x)\rvert$	Normalized Color at x.
$\boldsymbol{c}_r(x)$	$\int_{-\sigma_c}^{\sigma_c} \frac{1}{\sqrt{2\pi}\sigma_c}\ e^{-\frac{r^2}{2\sigma_c^2}}\ \boldsymbol{c}_n(x+r)\,dr$	Reference Color at x
$\mathcal{F}_L(r)$	$\frac{r+\sigma}{\sigma^2\sqrt{2\pi}}\ e^{-\frac{(r+\sigma)^2}{2\sigma^2}}$ $-\sigma < r < \sigma$ $\frac{r+\sigma}{\sigma_s^2\sqrt{2\pi}}\ e^{-\frac{(r+\sigma)^2}{2\sigma_s^2}}$ $-(\sigma+2\sigma_s) < r < -\sigma$ 0 otherwise	Left Half of Filter
$\mathcal{F}_R(r)$	$\mathcal{F}_L(-r)$	Right Half of Filter
$\mathcal{I}_L(x)$	$\int_{-(\sigma+\sigma_s)}^{\sigma} S_{\otimes}(\boldsymbol{c}_r(x), \boldsymbol{c}_n(x+r))\mathcal{F}_L(r)\,dr$	Inertia from Left Half
$\mathcal{I}_R(x)$	$\int_{-\sigma}^{\sigma+\sigma_s} S_{\otimes}(\boldsymbol{c}_r(x), \boldsymbol{c}_n(x+r))\mathcal{F}_R(r)\,dr$	Inertia from Right Half
$\mathcal{I}_\sigma(x)$	$\min(\mathcal{I}_L(x), \mathcal{I}_R(x))\ \frac{\sqrt{\sigma}}{(1+\mathcal{P}_{max}\frac{g(x)}{g_{max}})^2}$	Inertia at location x (Scale σ).
$\mathcal{I}(x)$	$\forall\sigma\ \max(\mathcal{I}_\sigma(x))$	Overall inertia at location x.
$\sigma(max)$	σ such that $\mathcal{I}_\sigma(x)$ is maximized	
$\mathcal{T}_L(x)$	0 if $r_c < \sigma(max)$ $r_c(\pi - \arccos(\frac{r_c-\sigma(max)}{r_c}))$ otherwise	Tolerated Length (Depends on radius of curvature r_c)

Table 1: Steps for Computing Directional Inertias and Tolerated Length. Note that the scale σ is *not* a free parameter.

Our approach uses two filters (see profile in Figure 8), each of which looks at one side of the ridge. The output of the combined filter is the minimum of the two responses. Each of the two parts of the filter is asymmetrical, reflecting the fact that we expect the object to be uniform (which explains each filter's large central lobe), and that we do not expect that a region of equal size be adjacent to the object (which explains each filter's small side lobe to accomodate for narrower adjacent regions). In other words, our ridge detector is designed to handle narrow valleys.

Handling steps and the extension to color are tricky because there is no clear notion of what is positive and what is negative in vector quantities. We solve this problem by adaptively defining a reference color at each point as the weighted average color over a small neighborhood of the point (about eight times smaller than the scale of the filter in the current implementation). Thus, this reference color will be different for different points in the image and scalar deviations from the reference color are computed as defined in section 3.

7 Filter Characteristics

This Section examines some interesting characteristics of our filter under noiseless and noisy operating conditions. We begin in Section 7.1 by deriving the filter's optimum scale response and its optimum scale map for noiseless ridge profiles, from which we see that both exhibit local output extrema at ridge centers. Next, we examine our filter's scale (Section 7.2) and spatial (Section 7.3) localization characteristics under varying degrees of noise. Scale localization measures the closeness in value between the optimum mask size at a ridge center and the actual width of the ridge. Spatial localization measures the closeness in position between the filter's peak response location and the actual ridge center. We shall see that both the filter's optimum scale and peak response location remain remarkably stable even at noticeably high noise levels. Our analysis will conclude with a comparison with Canny's ridge detector in Section 7.4 and experimental results in Section 11.

For simplicity, we shall perform our analysis on scalar ridge profiles instead of color ridge profiles. The extension to color is straightforward if we think of the *reference color* notion and the color similarity measure of equation 1 as a transformation that converts color ridge profiles into scalar ridge profiles.

We shall be using filter notations similar to those given in Table 6. In particular, σ denotes the main lobe's width (or scale), F_s denotes the filter's main lobe to side lobe width ratio, and $\mathcal{F}_L(r, \sigma_m, \sigma_s)$ a left-half filter with main lobe size σ_m, side lobe size $\sigma_s = \sigma_m / F_s$, and whose form is a normalized combination of two Gaussian first derivatives. At each point on a ridge profile, the filter outputs, by definition, the maximum response for mask pairs of all scales centered at that point.

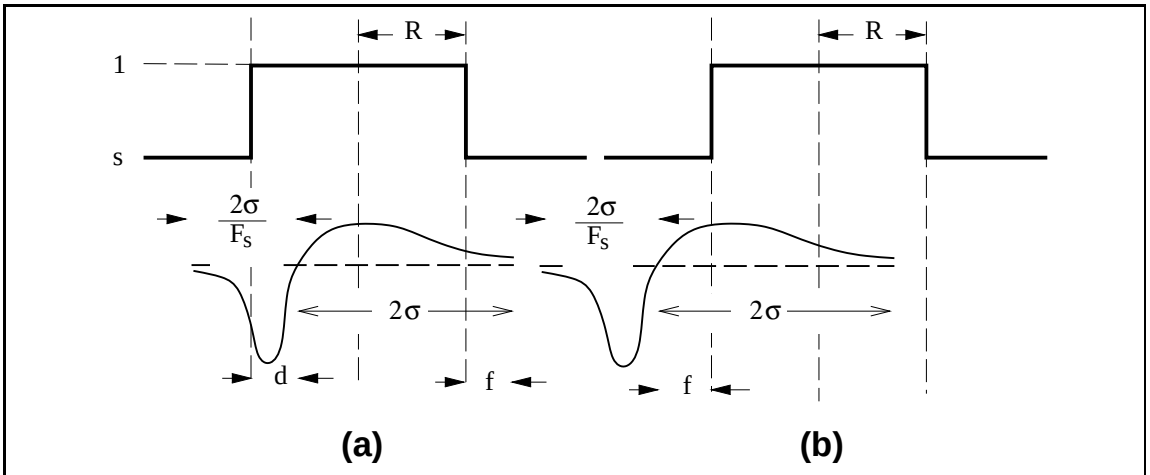


Figure 10: Half-mask configurations for computing the optimum scale ridge response of our filter. See text for explanation.

7.1 Filter Response and Optimum Scale

Let us first obtain the *single scale* filter response for the two half-mask configurations in Figure 10. Figure 10(a) shows an off-center left-half mask whose side lobe overlaps the ridge plateau by $0 \leq d \leq 2\sigma/F_s$ and whose main lobe partly falls off the right edge of the ridge plateau by $0 \leq f \leq 2\sigma$. The output in terms of mask dimensions and offset parameters is:

$$\begin{aligned} O_a(d,f) &= \int_{-(\sigma+\frac{2\sigma}{F_s})}^{-(\sigma+d)} s\mathcal{F}_L(r,\sigma,\frac{\sigma}{F_s})dr + \int_{-(\sigma+d)}^{\sigma-f} \mathcal{F}_L(r,\sigma,\frac{\sigma}{F_s})dr + \int_{\sigma-f}^{\sigma} s\mathcal{F}_L(r,\sigma,\frac{\sigma}{F_s})dr \\ &= \frac{1}{\sqrt{2\pi}}\Big[(F_s s-1)(e^{-2}-1) \\ &\quad -(1-s)\left(F_s(1-e^{-\frac{F_s^2 d^2}{2\sigma^2}}) + (e^{-\frac{(2\sigma-f)^2}{2\sigma^2}} - e^{-2})\right)\Big] \end{aligned} \tag{2}$$

A value of f greater than d indicates that the filter's main lobe (ie. its scale) is wider than the ridge and vice-versa. Notice that when $d = f = 0$, we have a perfectly centered mask whose main lobe width equals the ridge width, and whose output value is globally maximum.

Figure 10(b) shows another possible left-half mask configuration in which the main lobe partly falls outside the left edge of the ridge plateau by $0 \leq f \leq 2\sigma$. Its output is:

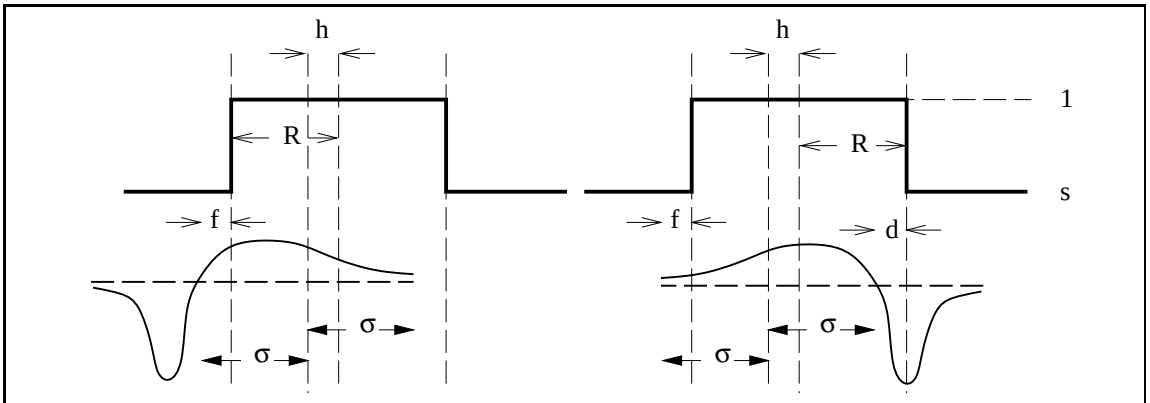


Figure 11: Mask-pair configuration for computing the all scales optimum ridge response of our filter. See text for explanation.

$$
\begin{aligned}
O_b(f) &= \int_{-(\sigma+\frac{2\sigma}{F_s})}^{-(\sigma-f)} s\mathcal{F}_L(r,\sigma,\frac{\sigma}{F_s})dr + \int_{-(\sigma-d)}^{\sigma} \mathcal{F}_L(r,\sigma,\frac{\sigma}{F_s})dr \\
&= \frac{1}{\sqrt{2\pi}}\left[(F_s s-1)(e^{-2}-1) - (1-s)(1-e^{-\frac{f^2}{2\sigma^2}})\right] \qquad (3)
\end{aligned}
$$

The equivalent right-half mask configurations are just mirror images of the two left-half mask configurations, and have similar *single scale* ridge response values.

Consider now the *all scales* optimum filter response of a mask pair, offset by h from the center of a ridge profile (see Figure 11). The values of d and f in the figure can be expressed in terms of the ridge radius (R), the filter size (σ) and the offset distance (h) as follows:

$$
\begin{aligned}
d &= R + h - \sigma \\
f &= \sigma + h - R
\end{aligned}
$$

Notice that the right-half mask configuration in Figure 11 is exactly the mirror image of the left-half mask configuration in Figure 10(a).

Because increasing σ causes f to increase which in turn causes the left-half mask output to decrease, while decreasing σ causes d to increase which in turn causes the right-half mask output to decrease, the *all scales* optimum filter response, $\texttt{Opt}(h,R)$, must therefore be from the scale, σ_o, whose left and right half response values are equal. Using the identities for

d and f above together with the half-mask response equations 2 and 3, we get, after some algebriac simplification:

$$\texttt{Opt}(h,R) = \frac{1}{\sqrt{2\pi}}\left[(F_s s - 1)(e^{-2} - 1) - (1-s)(1 - e^{-\frac{(\sigma_o + h - R)^2}{2\sigma_o^2}})\right] \tag{4}$$

where the *optimum scale*, σ_o, must satisfy the following equality:

$$F_s(1 - e^{-\frac{F_s^2(R+h-\sigma_o)^2}{2\sigma_o^2}}) + (e^{-\frac{(\sigma_o - h + R)^2}{2\sigma_o^2}} - e^{-2}) = (1 - e^{-\frac{(\sigma_o + h - R)^2}{2\sigma_o^2}}). \tag{5}$$

The following bounds for σ_o can be obtained:

$$\frac{R+h}{1 + \frac{\sqrt{2}}{F_s}\ln(\frac{F_s}{F_s - 1 - e^{-2}})} < \sigma_o < (R+h). \tag{6}$$

For our particular implementation, we have $F_s = 8$ which gives us: $0.9737(R+h) < \sigma_o < (R+h)$. Since $h \geq 0$, Equation 6 indicates that the optimum filter scale, σ_o, is a local *minimum* at ridge centers where $h = 0$.

To show that the *all scales* optimum filter response is indeed a local *maximum* at ridge centers, let us assume, using the inequality bounds in Equation 6, that $\sigma_o = k(R+h)$ for some fixed k in the range:

$$\frac{1}{1 + \frac{\sqrt{2}}{F_s}\ln(\frac{F_s}{F_s - 1 - e^{-2}})} < K < 1.$$

Equation 4 becomes:

$$\texttt{Opt}(h,R) = \frac{1}{\sqrt{2\pi}}\left[(F_s s - 1)(e^{-2} - 1) - (1-s)(1 - e^{-\frac{((1+k)h - (1-k)R)^2}{2k^2(R+h)^2}})\right]. \tag{7}$$

Differentiating the above equation with respect to h, we see that $\texttt{Opt}(h,R)$ indeed decreases with increasing h for values of h near 0.

7.2 Scale Localization

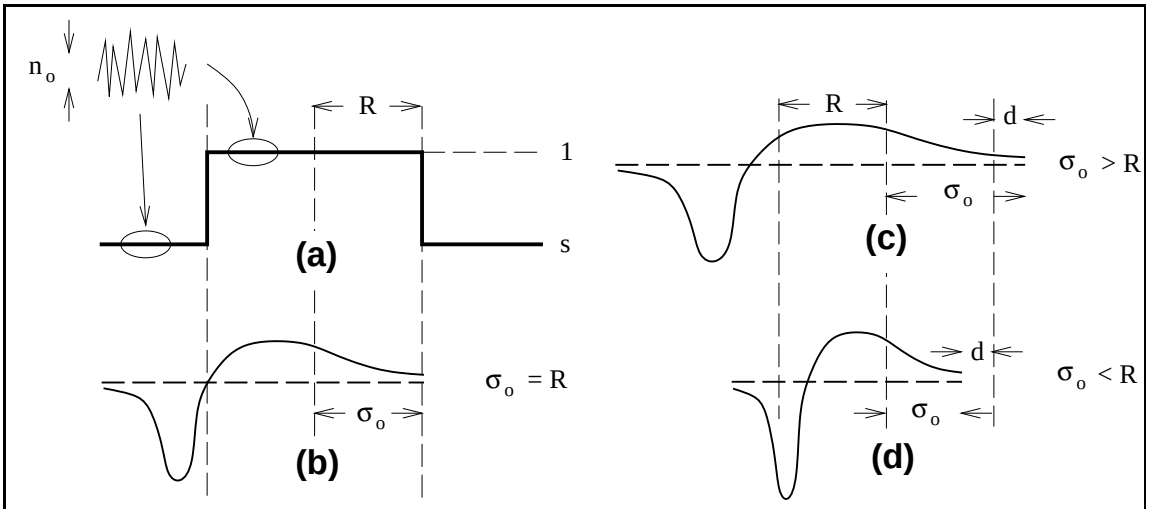


Figure 12: Mask configurations for scale localization analysis. (a) A radius R ridge profile with *noise to signal* ratio $n_o/(1-s)$. (b) A mask whose scale equals the ridge dimension. (c) A mask whose scale is larger than the ridge dimension. (d) A mask whose scale is smaller than the ridge dimension.

We shall approach the scale localization analysis as follows (see Figure 12(a)): Consider a radius R ridge profile whose *signal to noise ratio* is $(1-s)/n_o$, where $(1-s)$ is the height of the ridge signal and n_o^2 is the noise variance. Let $d = |R - \sigma_o|$ be the size difference between the ridge radius and the optimum filter scale at the ridge center. We want to obtain an estimate for the magnitude of d/R, which measures the *relative* error in scale due to noise.

Figures 12(b) (c) and (d) show three possible left-half mask configurations aligned with the ridge center. In the absence of noise (ie. if $n_o = 0$), their respective output values (O_s) are:

$$
\begin{aligned}
(\sigma = R) \qquad : O_s &= \int_{-(\sigma+\frac{2\sigma}{F_s})}^{-\sigma} s\mathcal{F}_L(r,\sigma,\frac{\sigma}{F_s})dr + \int_{-\sigma}^{\sigma} \mathcal{F}_L(r,\sigma,\frac{\sigma}{F_s})dr \\
&= \frac{1}{\sqrt{2\pi}}(1-e^{-2})(1-sF_s) \\
(\sigma = R+d) : O_s(d) &= \int_{-(\sigma+\frac{2\sigma}{F_s})}^{-(\sigma-d)} s\mathcal{F}_L(r,\sigma,\frac{\sigma}{F_s})dr + \int_{-(\sigma-d)}^{\sigma-d} \mathcal{F}_L(r,\sigma,\frac{\sigma}{F_s})dr \\
&\quad + \int_{\sigma-d}^{\sigma} s\mathcal{F}_L(r,\sigma,\frac{\sigma}{F_s})dr \\
&= \frac{1}{\sqrt{2\pi}}\Big[(1-e^{-2})(1-sF_s)
\end{aligned}
$$

$$+(1-s)(e^{-2}+e^{-\frac{d^2}{2(R+d)^2}}-e^{-2}e^{\frac{2d}{R+d}}e^{-\frac{d^2}{2(R+d)^2}}-1)\Bigg]$$

$$\begin{aligned}(\sigma = R-d):\ O_s(d) &= \int_{-(\sigma+\frac{2\sigma}{F_s})}^{-(\sigma+d)} s\mathcal{F}_L(r,\sigma,\frac{\sigma}{F_s})dr + \int_{-(\sigma+d)}^{\sigma}\mathcal{F}_L(r,\sigma,\frac{\sigma}{F_s})dr \\ &= \frac{1}{\sqrt{2\pi}}\left[(1-e^{-2})(1-sF_s)+(1-s)F_s(e^{-\frac{F_s^2d^2}{2(R-d)^2}}-1)\right] \qquad (8)\end{aligned}$$

Let us now compute O_n, the noise component of the filter output. Since the noise signal is white and zero mean, we have $\mathtt{E}[O_n]=0$, where $\mathtt{E}[x]$ stands for the expected value of x. For noise of variance n_o^2, the variance of O_n is:

$$\begin{aligned}\mathtt{Var}[O_n] &= \int_{-(\sigma+\frac{2\sigma}{F_s})}^{\sigma} n_o^2\mathcal{F}_L^2(r,\sigma,\frac{\sigma}{F_s})dr \approx \int_{-\infty}^{\infty} n_o^2\mathcal{F}_L^2(r,\sigma,\frac{\sigma}{F_s})dr \\ &= \frac{1+F_s}{8\sigma\sqrt{\pi}} \approx \frac{1+F_s}{8R\sqrt{\pi}}, \qquad (9)\end{aligned}$$

or equivalently, the standard deviation of O_n is:

$$\mathtt{SD}[O_n] = \sqrt{\frac{1+F_s}{8R\sqrt{\pi}}}. \qquad (10)$$

A very loose upper bound for d/R can be obtained by finding d, such that the noiseless response for a size $\sigma = R+d$ (or size $\sigma = R-d$) mask is within one noise output standard deviation of the optimum scale response (ie. the response for a mask of size $\sigma_o = R$). We examine first, the case when $\sigma = R+d$. Subtracting O_s for $\sigma = R$ from $O_s(d)$ for $\sigma = R+d$ (both from the series of equations 8) and equating the difference with $\mathtt{SD}[O_n]$, we get:

$$(1-s)(1-e^{-2}+e^{-\frac{d^2}{2(R+d)^2}}-e^{-2}e^{\frac{2d}{R+d}}e^{-\frac{d^2}{2(R+d)^2}}) = \sqrt{\frac{1+F_s}{8R\sqrt{\pi}}},$$

which, after some algebra and simplifying approximations, becomes:

$$\begin{aligned}d/R &\approx \frac{\sqrt{2K}}{1-\sqrt{2K}} \qquad \left(0 \le \frac{n_o}{1-s} < (1-e^{-2})(1-e^{-\frac{1}{2}})\sqrt{\frac{8R\sqrt{\pi}}{1+F_s}}\right) \\ &\text{where :} \qquad K = -\ln\left(1-\frac{n_o}{1-s}\frac{1}{1-e^{-2}}\sqrt{\frac{1+F_s}{8R\sqrt{\pi}}}\right). \qquad (11)\end{aligned}$$

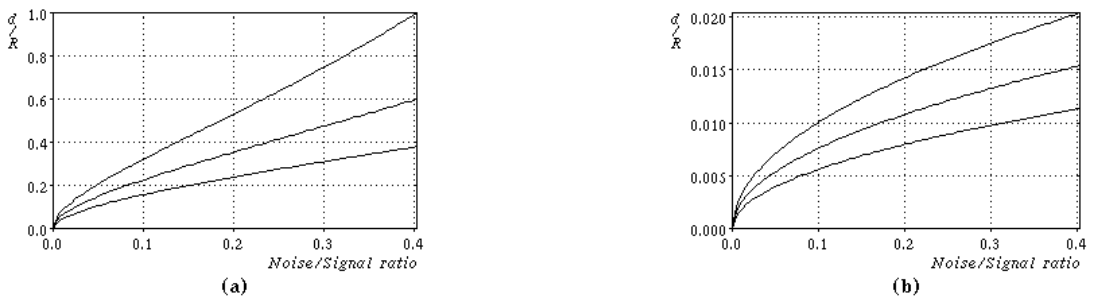


Figure 13: Relative scale error (d/R) as a function of *noise to signal ratio* ($n_o/(1-s)$) for *(a) Equation 11 where* $\sigma_o > R$, and *(b) Equation 12 where* $\sigma_o < R$. For both graphs, $F_s = 8$, top curve is for $R = 10$, middle curve is for $R = 30$ and bottom curve is for $R = 100$.

Figure 13(a) graphs d/R as a function of the *noise to signal ratio* $n_o/(1-s)$. We remind the reader that our derivation is in fact a probabilistic upper bound for d/R. For d/R to exceed the bound, the $\sigma = R + d$ filter must actually produce a combined signal and noise response, greater than that of all the other filters with sizes from $\sigma = R$ to $\sigma = R + d$.

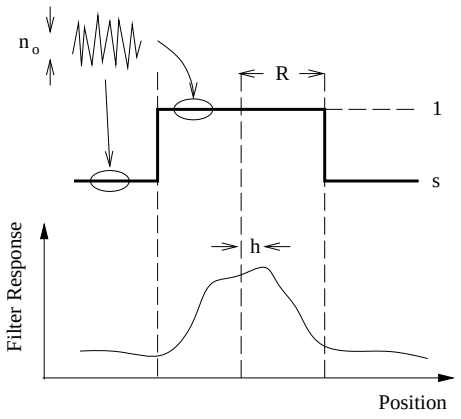
A similar analysis for the $\sigma = R - d$ case yields (see Figure 13(b) for plot):

$$d/R \approx \frac{\sqrt{2K}}{F_s + \sqrt{2K}}$$
$$\text{where :} \qquad K = -\ln\left(1 - \frac{n_o}{1-s}\sqrt{\frac{1+F_s}{8F_s^2 R\sqrt{\pi}}}\right). \tag{12}$$

7.3 Spatial Localization

Consider the radius R ridge in Figure 14 whose signal to noise ratio is $(1-s)/n_o$. As before, $(1-s)$ is the height of the ridge signal and n_o^2 is the noise variance. Let h be the distance between the actual ridge center and the peak location of the filter's *all scales* ridge response. Our goal is to establish some magnitude bound for h/R that can be brought about by the given noise level.

To make our analysis feasible, let us assume, using Equation 6, that the optimum filter scale at distance h from the ridge center is $\sigma_o = R + h$. Notice that for our typical values of F_s, the uncertainty bounds for σ_o is relatively small. The optimum scale filter output without noise is therefore:



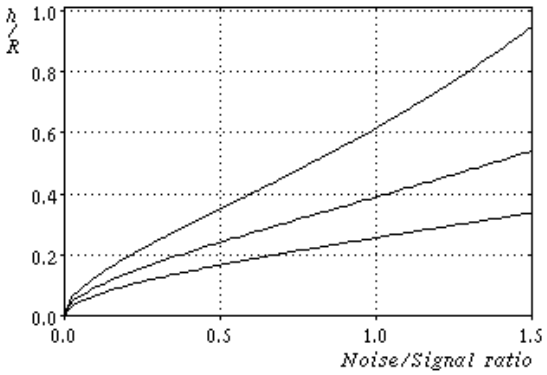


Figure 14: **Left:** Mask configurations for scale localization analysis. An all scales filter response for a radius R ridge profile with *noise to signal* ratio $n_o/(1-s)$. h is the distance between the actual ridge center and the filter response peak location. **Right:** Relative spatial error (h/R) as a function of *noise to signal ratio* $(n_o/(1-s))$, where $F_s = 8$, top curve is for $R = 10$, middle curve is for $R = 30$ and bottom curve is for $R = 100$. See Equation 15.

$$\texttt{Opt}(h,R) \approx \frac{1}{\sqrt{2\pi}}\left[(F_s s-1)(e^{-2}-1)-(1-s)(1-e^{-\frac{4h^2}{2(R+h)^2}})\right], \tag{13}$$

and the difference in value between the above and the noiseless optimum scale output at ridge center is:

$$\texttt{Opt}(0,R)-\texttt{Opt}(h,R) \approx (1-s)(1-e^{-\frac{4h^2}{2(R+h)^2}}). \tag{14}$$

As in the scale localization case, we obtain an estimate for h/R by finding h such that the difference in Equation 14 equals one noise output standard deviation of the optimum scale filter at ridge center (see Equation 10). We get:

$$(1-s)(1-e^{-\frac{4h^2}{2(R+h)^2}}) = n_o\sqrt{\frac{1+F_s}{8R\sqrt{\pi}}},$$

which eventually yields (see Figure 14 for plot):

$$h/R = \frac{\sqrt{K}}{\sqrt{2}-\sqrt{K}} \qquad (0 \le \frac{n_o}{1-s} < (1-e^{-2})\sqrt{\frac{8R\sqrt{\pi}}{1+F_s}})$$

$$\text{where :} \qquad K = -\ln\left(1 - \frac{n_o}{1-s}\sqrt{\frac{1+F_s}{8R\sqrt{\pi}}}\right). \tag{15}$$

7.4 Scale and Spatial Localization Characteristics of the Canny Ridge Operator

We compared our filter's scale and spatial localization characteristics with those of a Canny ridge operator. This is a relevant comparison because the Canny ridge operator was designed to be optimal for simple ridge profiles (see [Canny 1985] for details on the optimality criterion). The normalized form of Canny's ridge detector can be approximated by the shape of a scaled Gaussian second derivative:

$$\mathcal{C}(r,\sigma) = \frac{1}{\sqrt{2\pi}\sigma^3}(\sigma^2 - r^2)e^{-\frac{r^2}{2\sigma^2}}. \tag{16}$$

We begin with scale localization. For a noiseless ridge profile with radius R and height $(1-s)$, the optimum scale ($\sigma = R$) Canny filter response at the ridge center is:

$$O_s(\sigma = R) = \sqrt{\frac{2}{\pi}}(1-s)e^{-\frac{1}{2}}. \tag{17}$$

Similarly, the ridge center filter response for a mis-matched Canny mask ($\sigma = R + d$) is:

$$O_s(\sigma = R+d) = \sqrt{\frac{2}{\pi}}\frac{R}{R+d}(1-s)e^{-\frac{R^2}{2(R+d)^2}},$$

where the scale difference, d, can be either positive or negative in value.

We want an estimate of d/R in terms of the noise to signal ratio. Consider now the effect of white Gaussian noise (zero mean and variance n_o^2) on the optimum scale Canny filter response. The noise output standard deviation is:

$$\mathtt{SD[O_n]} = \sqrt{\int_{-\infty}^{\infty} n_o^2 \mathcal{C}^2(r, \sigma = R)dr}$$

$$= \quad n_o\sqrt{\frac{3}{8R\sqrt{\pi}}}. \tag{18}$$

Performing the same scale localization steps as we did for our filter, we get:

$$n_o\sqrt{\frac{3}{8R\sqrt{\pi}}} = \sqrt{\frac{2}{\pi}}e^{-\frac{1}{2}}(1-s) - \sqrt{\frac{2}{\pi}}\frac{R}{R+d}e^{-\frac{R^2}{(R+d)^2}}(1-s),$$

which reduces to the following equation that implicitly relates d/R to $\frac{n_o}{1-s}$:

$$\frac{n_o}{1-s} = \sqrt{\frac{16R}{3\sqrt{\pi}}}\left[e^{-\frac{1}{2}} - \frac{R}{R+d}e^{-\frac{R^2}{2(R+d)^2}}\right]. \tag{19}$$

For spatial localization, we want an estimate of h/R in terms of $\frac{n_o}{1-s}$, where h is the distance between the actual ridge center and the *all scales* Canny operator peak output location. At distance h from the ridge center, the optimum Canny mask scale (σ_o) is bounded by:

$$\sqrt{\mathrm{R^2 + h^2 - 2Rh}\frac{1-\mathrm{e}^{-\frac{4\mathrm{Rh}}{2(\mathrm{R-h})^2}}}{1+\mathrm{e}^{-\frac{4\mathrm{Rh}}{2(\mathrm{R-h})^2}}}} \leq \sigma_\mathrm{o} \leq \sqrt{\mathrm{R^2 + h^2 - 2Rh}\frac{1-\mathrm{e}^{-\frac{4\mathrm{Rh}}{2(\mathrm{R+h})^2}}}{1+\mathrm{e}^{-\frac{4\mathrm{Rh}}{2(\mathrm{R+h})^2}}}},$$

and the noiseless optimum scale filter response is:

$$O_s(h) = \frac{2}{\sqrt{2\pi}\sigma_o}(1-s)e^{-\frac{R^2+h^2}{2\sigma_o^2}}\left[R\cosh(\frac{Rh}{\sigma_o^2}) - h\sinh(\frac{Rh}{\sigma_o^2})\right].$$

Setting $O_s(0) - O_s(h) = \mathtt{SD}[O_n]$, we arrive at the following implicit equation relating h/R and $n_o/(1-s)$:

$$\frac{n_o}{1-s} \approx \sqrt{\frac{4R}{3\sqrt{\pi}}}\left[e^{-\frac{1}{2}} - \frac{1}{\sigma_o}e^{-\frac{R^2+h^2}{2\sigma_o^2}}\left(R\cosh(\frac{Rh}{\sigma_o^2}) - h\sinh(\frac{Rh}{\sigma_o^2})\right)\right], \tag{20}$$

where $\sigma_o \approx \sqrt{R^2 + h^2 - 2Rh(1-e^{-\frac{4Rh}{2R^2}})/(1+e^{-\frac{4Rh}{2R^2}})}$ (valid for small h/R values).

We see from Figures 15 and 16 that at typical F_s ratios, our filter's scale and spatial localization characteristics are comparable to those of the Canny ridge operator.

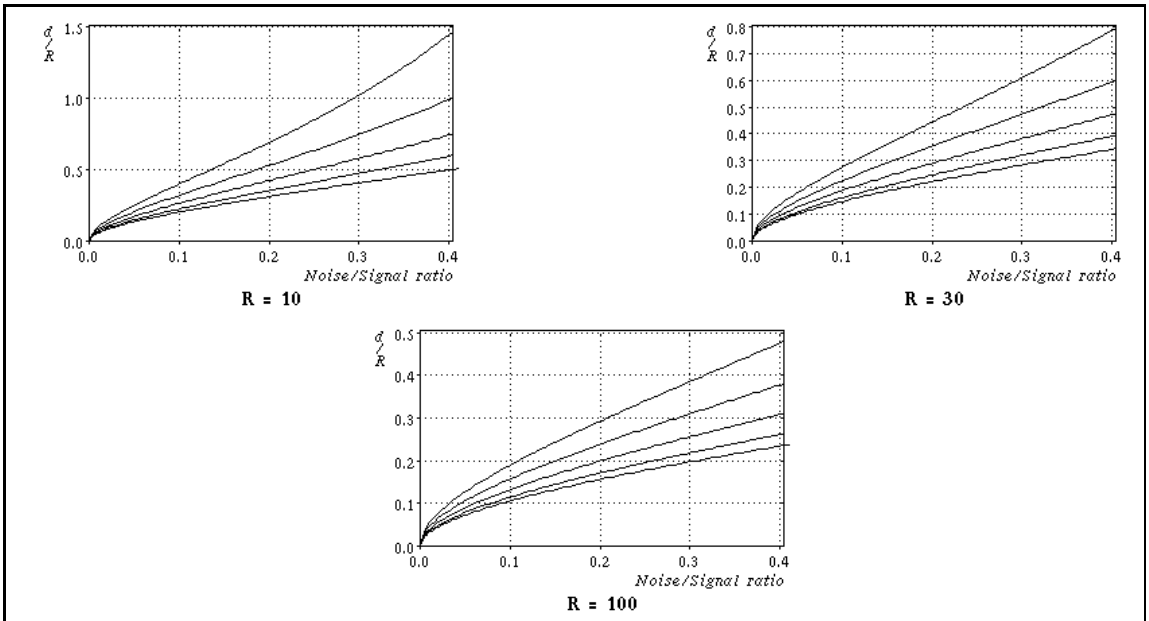


Figure 15: Comparison of relative scale error (d/R) as a function of *noise to signal ratio* ($n_o/(1-s)$) between our filter ($\sigma > R$ case) and the Canny ridge filter. See Equations 11 and 19. *Top Left:* $R = 10$. *Top Right:* $R = 30$. *Bottom:* $R = 100$. For each graph, curves from top to bottom are those of: $F_s = 16$, $F_s = 8$, $F_s = 4$, $F_s = 2$, and Canny.

8 Finding 2D Skeletons Using Directional 1D Ridge Detectors

The scheme that we present in this paper is an extension of Curved Inertia Frames (CIF), a brightness-based segmentation scheme presented in [Subirana-Vilanova 1990], which in turn is an extension of an *edge*-based perceptual organization scheme presented in the same paper. We choose this scheme for two reasons, first it is the only existing scheme that can compute global regions directly on the image without imposing a three-dimensional representation of the data. Second, we have been able to overcome a number of problems in the scheme making it is useful for a large class of images.

[Subirana-Vilanova 1990]'s scheme (and ours) proceeds in three stages. In the first one, it computes two local measures at each point p for a number of orientations θ: the *inertia value* $\mathcal{I}(p,\theta)$ and the *tolerated length* $\mathcal{T}(p,\theta)$. These two local values are based on the output of elongated gabor filters and are used to associate a *saliency measure* to *each* curve $C(t)$ in the image plane as defined in equation 21. Were the curve is assumed to be parameterized between 0 and L. $\mathcal{I}(l)$ ($\mathcal{T}(t)$) is the inertia value (tolerated length) at the point with parameter l and with the orientation of the curve at that point, and ρ and α are suitable constants.

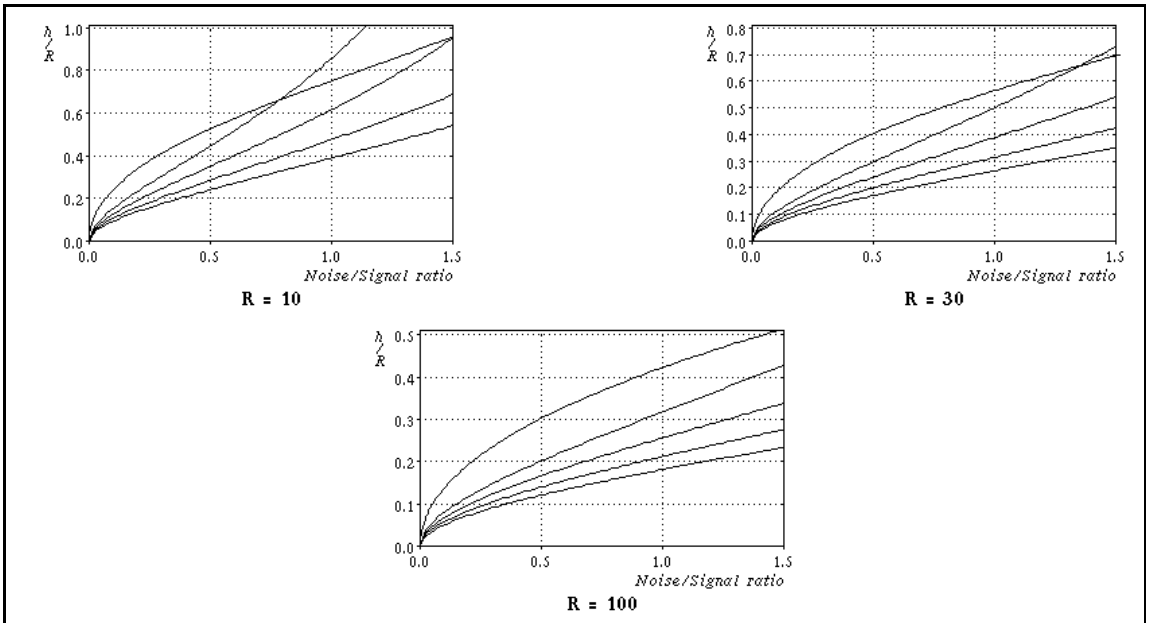


Figure 16: Comparison of relative spatial error (h/R) as a function of *noise to signal ratio* $(n_o/(1-s))$ between our filter and the Canny ridge filter. See Equations 15 and 20. *Top Left:* $R = 10$. *Top Right:* $R = 30$. *Bottom:* $R = 100$. For each graph, the Canny curve is the top curve between $n_o/(1-s) = 0$ and $n_o/(1-s) = 0.5$. The other curves from top to bottom are for: $F_s = 16$, $F_s = 8$, $F_s = 4$ and $F_s = 2$.

$$S_L = \int_0^L \mathcal{I}(l)\rho^{\int_0^l \frac{1}{\alpha \mathcal{T}(t)}dt}dl \tag{21}$$

In the second stage, the scheme computes the skeleton which yields the maximum saliency using an extension of the network introduced by [Shashua and Ullman 1988]. In fact, the form of equation 21 closely matches what the network can compute. The inertia value and the tolerated length can be used in the second stage using other schemes such as [Kass, Witkin and Terzopoulos 88], [Zucker, Dobbins and Iverson 89], and [Pizer, Burbeck, and Coggins 1993].

The scheme favors curves which are long, smooth (according to the associated tolerated length values) and central to the shape (i.e. which have high inertia values). This second stage yields the *skeleton sketch* a representation of the potential skeletons in the image. See [Subirana-Vilanova 1990], [Subirana-Vilanova 1991] for more details.

In the third stage, the scheme computes a succession of individual curves (or skeletons) and the corresponding perceptual groups by growing outward from the skeletons.

In this section we will derive a class of dynamic programming algorithms that find curves in an arbitrary graph that maximize a certain quantity. In the next sections we will apply these algorithms to finding long and smooth ridges in the inertia surfaces, which are the output of our one dimensional filter when applied at different orientations. [Mahoney 1987] showed that long and smooth curves in binary images are salient in human perception even if they have multiple gaps and in the presence of other curves. [Sha'ashua and Ullman 1988] devised a saliency measure and a dynamic programming algorithm that can find such salient curves in a binary image (see also [Ullman 1976]). We build on their work and show how their ideas can be extended to deal with arbitrary surfaces. In this section we will examine their computation in a way geared at demonstrating that the kind of saliency measures that can be computed with the network is very limited. The actual proof of this will be given in Section 10.

We define a *directed graph with properties* $G = (V, E, P_E, P_J)$ as a graph with a set of *vertices* $V = \{v_i\}$; a set of *edges* $E = \{e_{i,j} = (v_i, v_j) \mid v_i, v_j \in V\}$; a function $P_E : E \rightarrow \Re$ that assigns a vector $\boldsymbol{p}_e$ of *properties* to each edge; and a function $P_J : J \rightarrow \Re$ that assigns a vector $\boldsymbol{p}_j$ of *properties* to each *junction* where a junction is a pair of adjacent edges (i.e. any pair of edges that share a vertex) and J is the set of all junctions. We will refer to a curve in the graph as a sequence of connected edges. We assume that we have a *saliency function* S that associates a positive integer $S(C)$ with each curve C in the graph. This integer is the *saliency* or *saliency value* of the curve. The saliency of a curve will be defined in terms of the properties of the elements (vertices, edges and junctions) of the curve.

Our problem is to find a computation that finds for every point and each of its connecting edges, the most salient curve starting at that point with that edge. This includes defining a saliency function and a computation that will find the salient curves for that function. The applications that will be shown here work with a 2 dimensional grid. The vertices are the points in the grid and the edges the elements that connect the different points in the grid. The junctions will be used to include in the saliency function *properties* of the shape of the curve such as curvature.

The computation will be performed in a locally connected parallel network with a processor $pe_{i,j}$ for every edge $e_{i,j}$. The processors corresponding to the incoming edges of a given vertex will be connected to those corresponding to the connecting edges at that vertex. We will design the computation so that we know at iteration n what is the saliency of the most salient curve of size n for every edge. This provides a constraint in the invariant of the algorithm that we are seeking that will guide us to the final algorithm. In order for the computation to have some computing power each processor $pe_{i,j}$ must have at least *one* state variable that we will denote as $s_{i,j}$. Since we want to know the saliency of the most salient curve of length n starting with any given edge, we will assume that, at iteration n, $s_{i,j}$ contains that value for that edge. Observe that having only one variable looks like a big restriction, however, we show in Section 10 that *allowing more state variables does not add any power to the possible saliency functions that can be computed with this*

network. Since the saliency of a curve is defined only by the properties of the elements in the curve, it cannot be influenced by properties of elements outside the curve. Therefore the computation to be performed can be expressed as:

$$s_{i,j}(n+1) = \text{MAX}\{\mathcal{F}(n+1, \boldsymbol{p}_e, \boldsymbol{p}_j, s_{i,j}(n), s_{j,k}(n)) \mid \quad (j,k) \in E\}$$

$$s_{i,j}(0) = \mathcal{F}(0, \boldsymbol{p}_e, \boldsymbol{p}_j, 0, 0) \tag{22}$$

where $\mathcal{F}$ is the function that will be computed in every iteration and that will lead to the computed saliency. Observe that given $\mathcal{F}$, the saliency value of any curve can be found by applying $\mathcal{F}$ recursively on the elements of the curve.

We are now interested in what types of saliency functions S we can use and what type of functions $\mathcal{F}$ are needed to compute them such that the value obtained in the computation is the maximum for the resulting saliency measure S. Using contradiction and induction we conclude that a function $\mathcal{F}$ will compute the most salient curve for all possible graphs if and only if it is monotonically increasing in its last argument. That is, if and only if:

$$\forall \boldsymbol{p}, x, y \;\; x < y \;\; \longrightarrow \mathcal{F}(\boldsymbol{p}, x) < \mathcal{F}(\boldsymbol{p}, y), \tag{23}$$

where $\boldsymbol{p}$ is used to abbreviate the first four arguments of $\mathcal{F}$.

What type of functions $\mathcal{F}$ satisfy this condition? We expect them to behave freely as $\boldsymbol{p}$ varies. And when $s_{j,k}$ varies, we expect $\mathcal{F}$ to change in the same direction with an amount that depends on $\boldsymbol{p}$. A simple way to fulfill this condition is with the following function:

$$\mathcal{F}(\boldsymbol{p}, x) = f(\boldsymbol{p}) + g(x) * h(\boldsymbol{p}) \tag{24}$$

where f, g and h are positive functions and g is monotonically increasing.

We now know what type of function $\mathcal{F}$ we should use but we do not know what type of saliency measures we can compute. Let us start by looking at the saliency S_i that we would compute for a curve of length i. For simplicity we assume that g is the identity function:

- **Iter. 1:** $S_1 = f(\boldsymbol{p}_{1,2})$

- **Iter. 2:** $S_2 = S_1 + f(\boldsymbol{p}_{2,3}) * h(\boldsymbol{p}_{1,2})$
- **Iter. 3:** $S_3 = S_2 + f(\boldsymbol{p}_{3,4}) * h(\boldsymbol{p}_{1,2}) * h(\boldsymbol{p}_{2,3})$
- **Iter. 4:** $S_4 = S_3 + f(\boldsymbol{p}_{4,5}) * h(\boldsymbol{p}_{1,2}) * h(\boldsymbol{p}_{2,3}) * h(\boldsymbol{p}_{3,4})$

  ...
- **Iter. i:** $S_i = S_{i-1} + f(\boldsymbol{p}_{i,i-1}) * \prod_{k=1}^{k=i-1} h(\boldsymbol{p}_{k,k+1}) = \sum_{l=1}^{l=i} f(\boldsymbol{p}_{l,l-1}) * \prod_{k=1}^{k=l-1} h(\boldsymbol{p}_{k,k+1})$.

At step n, the network will know about the most salient curve of length n starting from any edge. Recovering the most salient curve from a given point can be done by tracing the links chosen by the processors (from Equation 22).

9 Finding Long And Smooth Ridges

In this section, we will show how the network defined in the previous section can be used to find frames of reference using the *inertia surfaces* and the *tolerated length* as defined in the previous sections. The *directed graph with properties* that defines the network has one vertex for every pixel in the image and one edge connecting it to each of its neighbors thus yielding a locally connected parallel network. This results in a network that has eight orientations per pixel. The number of orientations per pixel can be increased to improve the accuracy of the output.

The value computed is the sum of the $f(\boldsymbol{p}_{i,j})$'s along the curve weighted by the product of the $h(\boldsymbol{p}_{i,j})$'s. Using $0 \leq h \leq 1$ we can ensure that the total saliency will be smaller than the sum of the f's. One way of achieving this is by using $h = 1/k$ or $h = \exp(-k)$ and restricting k to be larger than 1. The f's will then be a quantity to be maximized and the k's a quantity to be minimized along the curve. In our skeleton network (presented in the next section), f will be the inertia measure and k will depend on the tolerated length and will account for the shape of the curve so that the saliency of a curve is the sum of the inertia values along a curve weighted by a number that depends on the overall smoothness of the curve. In particular, the functions f, g and h (see Equation 24) are defined as:

- $f(\boldsymbol{p}) = f(\boldsymbol{p}_e) = \mathcal{I}(x)$,
- $g(x) = x$
- and $h(\boldsymbol{p}) = h(\boldsymbol{p}_j) = \rho^{\frac{l_{emt}}{\alpha \mathcal{T}(\boldsymbol{p}_j(\boldsymbol{x}))}}$.

α, which we call the *circle constant*, scales the tolerated length, and it was set to 4 in the current implementation (because $4\,radius\pi/2$ is the length of the perimeter of a circle). ρ, which we call the *penetration factor*, was set to 0.5 (so that inertia values "half a circle" away get factored down by 0.5). And l_{emt} is the length of the corresponding element. Also, $s_{i,j}(0) = 0$ (because the *saliency* of a skeleton of length 0 should be 0).

With this definition the *saliency value* assigned to a curve of length L is:

$$S_L = \sum_{l=1}^{l=L} \mathcal{I}(\ \boldsymbol{p}_{l,l-1}) \prod_{k=1}^{k=l-1} \rho^{\frac{l_{emt}}{\alpha \mathcal{T}(\boldsymbol{p}_k)}} = \sum_{l=1}^{l=i} \mathcal{I}(\boldsymbol{p}_{l,l-1}) \rho^{\sum_{k=1}^{k=l-1} \frac{l_{emt}}{\alpha \mathcal{T}(\boldsymbol{p}_k)}},$$

which is an approximation of the continuous value given in Equation 25 below. S_L is the saliency of a parameterized curve $C(u)$, and $\mathcal{I}(u)$ and $\mathcal{T}(u)$ are the inertia value and the tolerated length respectively at point u of the curve.

$$S_L = \int_0^L \mathcal{I}(l) \rho^{\int_0^l \frac{1}{\alpha \mathcal{T}(t)} dt} dl \tag{25}$$

The obtained measure favors curves that lie in large and central areas of the shape and that have a low overall internal curvature. The measure is bounded by the area of the shape; e.g. a straight symmetry axis of a convex shape will have a saliency equal to the area of the shape. In the next section we will present some results showing the robustness of the scheme in the presence of noisy shapes.

Observe that if the tolerated length $\mathcal{T}(t)$ at one point $C(t)$ is small then $\int_0^l \frac{1}{\alpha \mathcal{T}(t)} dt$ is large so that $\rho^{\int_0^l \frac{1}{\alpha \mathcal{T}(t)} dt} dl$ becomes very small (since $\rho < 1$) and so does the saliency for the curve S_L. Thus, a small α or ρ penalize curvature favoring smoother curves.

10 Limitations of the Dynamic Programming Approach

In this section we show that the set of possible saliency measures that can be computed with the network defined in the previous sections is limited.

Proposition 1 *The use of more than one state variable in the saliency network defined in the previous sections does not increase the set of possible saliency functions that can be computed with the network.*

Proof: The notation used in the proof will be the one used in the previous sections. We will do the proof for the case of two state variables, the generalization of the proof to more state variables follows naturally. Assume then, that each edge has a saliency state variable $s_{i,j}$ and an auxiliary state variable $a_{i,j}$ and two functions to update the state variables: $s_{i,j}(n+1) = MAX_k \mathcal{F}(\boldsymbol{p}, s_{j,k}(n), a_{j,k}(n))$ and $a_{i,j}(n+1) = \mathcal{G}(\boldsymbol{p}, s_{j,k}(n), a_{j,k}(n))$. We will show that for any pair of functions $\mathcal{F}$ and $\mathcal{G}$ either they can be reduced to one function or there is a network for which they do not compute the optimal curves.

If $\mathcal{F}$ does not depend on its last argument $a_{j,k}$ then the decision of what is the most salient curve is not affected by the introduction of more state variables so we can do without them. Observe that we might still use the state variables to compute additional properties of the most salient curve without affecting the actual shape of the computed curve.

If $\mathcal{F}$ does depend on its last argument then there exists some $\boldsymbol{p}$, x, y and $w \in \Re$ such that: $\mathcal{F}(\boldsymbol{p}, y, x) < \mathcal{F}(\boldsymbol{p}, y, w)$. Assuming continuity this implies that there exists some $\epsilon > 0$ such that: $\mathcal{F}(\boldsymbol{p},\ y - \epsilon\ , x) < \mathcal{F}(\boldsymbol{p}, y, w)$. Assume now two curves of length n starting from the same edge $e_{i,j}$ such that $s1_{i,j}(n) = y$, $a1_{i,j}(n) = x$, $s2_{i,j}(n) = y - \epsilon$ and $a2_{i,j}(n) = y$. If the algorithm where correct at iteration n it would have computed the values $s1_{i,j}(n) = y$, $a1_{i,j}(n) = x$ for the variables $s_{i,j}$ and $a_{i,j}$. But then at iteration n+1 the saliency value computed for an edge $e_{h,i}$ would be $s_{h,i} = \mathcal{F}(\boldsymbol{p},\ y - \epsilon\ , x)$ instead of $\mathcal{F}(\boldsymbol{p}, y, w)$ that corresponds to a curve with a higher saliency value. $\Box$.

11 Results

We have tested our scheme (filter + network) extensively, Figures 17 and 18 show that our filter produces sharper and more stable ridge responses than the second derivative of a gaussian filter, even when working with the notion of reference colors for color ridge profiles. First, our filter localizes all the ridges for a single ridge, for multiple or step ridges and for noisy ridges. The second derivative of the gaussian instead fails under the presence of multiple or step ridges. Second, the scale chosen by our operator matches the underlying data closely while the scale chosen by the second derivative of the gaussian does not match the underlying data (see Figures in Section 7). This is important because the scale is necessary to compute the Tolerated Length which is used in the second stage of our scheme to find the Curved Inertia Frames of the image. And third, our filter does not respond to edges while the second derivative of the gaussian does.

In the previous paragraph, we have discussed the one-dimensional version of our filter. The same filter can be used as a directional ridge operator for two-dimensional images. Figure 21 shows the directional output (aka inertia surfaces) of our filter on four images. The two-dimensional version of the filter can be used with different degrees of elongation. In our experiments we used one pixel width to study the worst possible scenario. An elongated

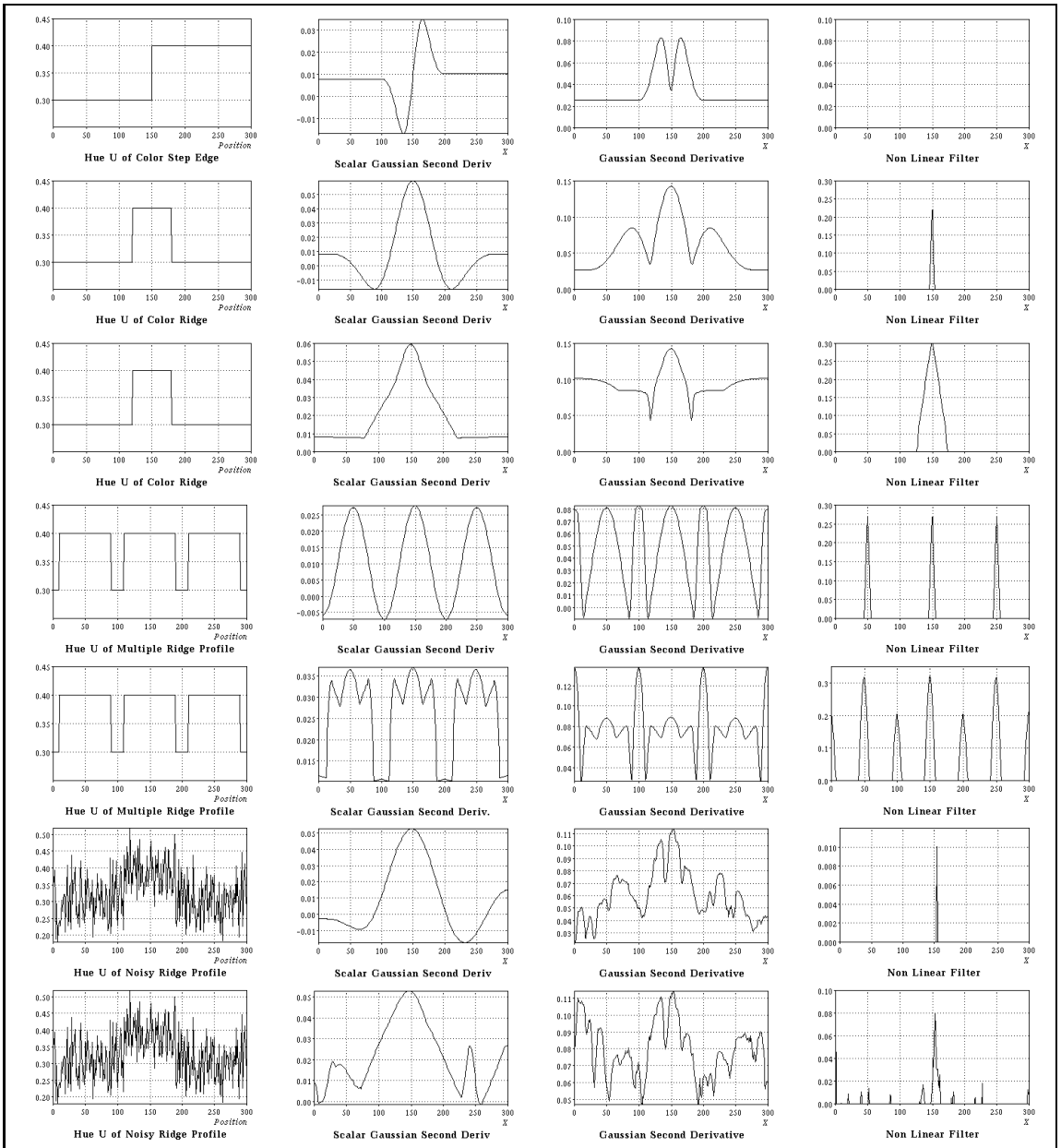


Figure 17: *First column:* Different input signals. *Second column:* Output given by second derivative of the gaussian. *Third column:* Output given by second derivative of the gaussian using reference color. *Fourth column:* Output given by our ridge detector. The *First, Second, Fourth and Sixth rows* are results of a single scale filter application where σ is tuned to the size of the largest ridge. The *Third, Fifth and Seventh rows* are results of a multiple scale filter application. Note that no scale parameter is involved in any multiple-scale case.

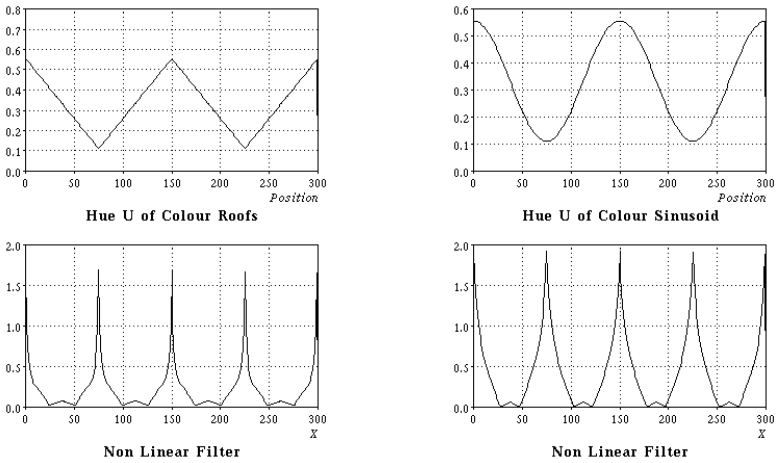


Figure 18: Comparing multiple scale filter responses two color profiles. *Top:* Hue U channel of roof and sinusoid color profiles. *Bottom:* Multi-scale output given by *color* convolution of our non-linear mask with the color profiles. Even though our filter was designed to detect flat regions, it can also detect other type of regions.

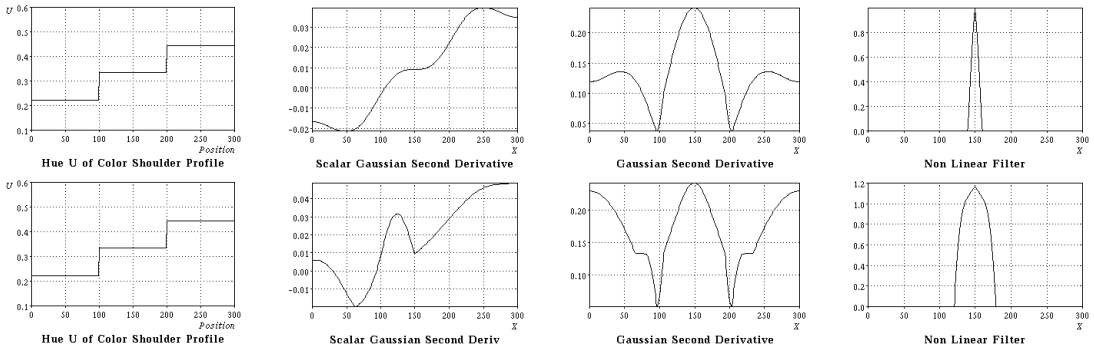


Figure 19: *First column:* Multiple step input signal. *Second column:* Output given by second derivative of the gaussian. *Third column:* Output given by second derivative of the gaussian using reference color. *Fourth column:* Output given by our ridge detector. The *first row* shows results of a single scale filter application where σ is tuned to the size of the largest ridge. The *second row* shows results of a multiple scale filter application. Note that no scale parameter is involved in multiple-scale case.

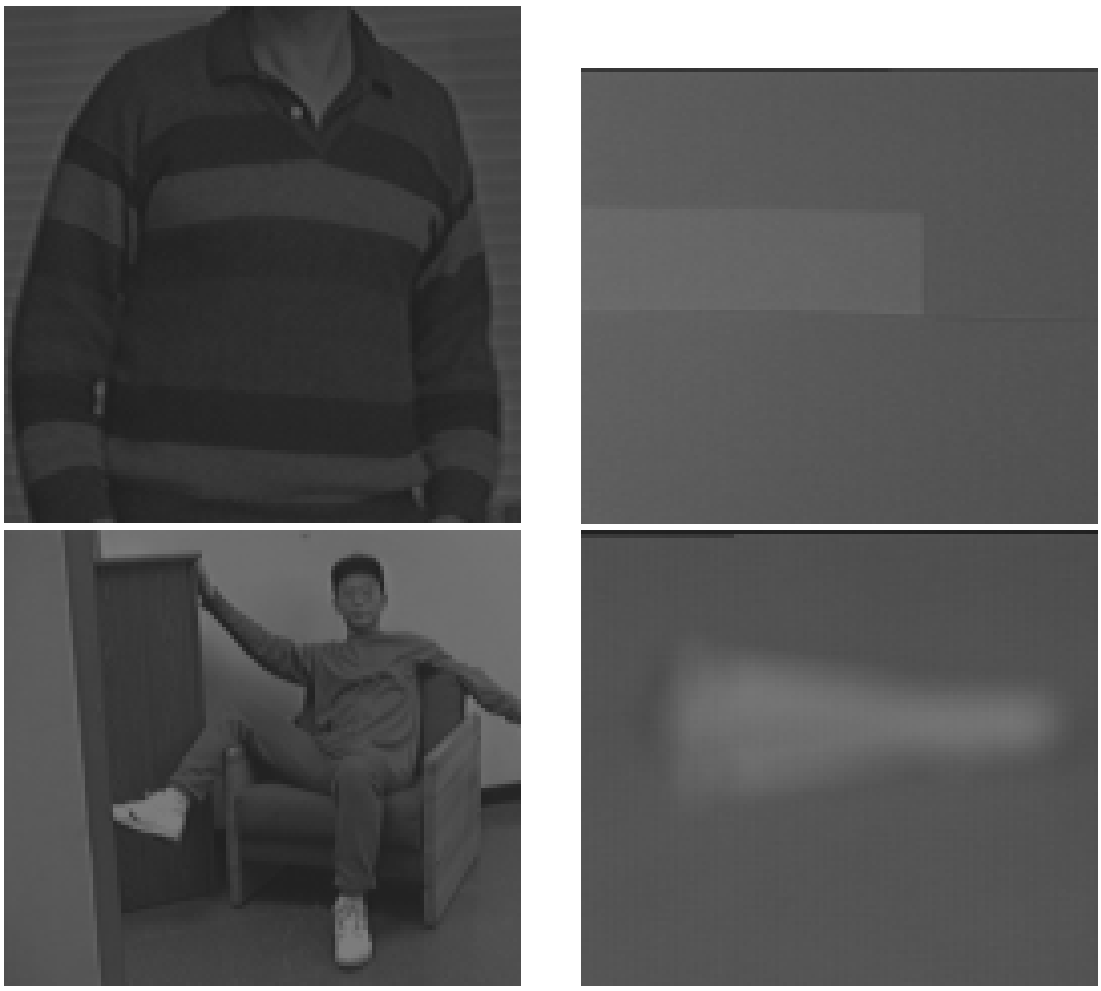


Figure 20: Four images: Sweater image, Ribbons image, Person image and Blob image. See inertia surfaces for these images in Figure 18 and the Canny edges at different scales for the Person and Blob image in Figure 5. Note that our scheme recovers the Person and blob at the right scale, without the need of specifying the scale.

filter would smooth existing noise; however, large scales are not good because they smooth the response near discontinuities and in curved areas of the shape (this can be overcome by using curved filters [Malik and Gigus 1991]).

The inertia surfaces and the tolerated length are the output of the first stage of our scheme. In the second stage we use these to compute the Curved Inertia Frames (see [Subirana-Vilanova 1990]) as shown in Figures 23, 24, 25, 26, and 27. These skeleton representations are used to grow the corresponding regions by a simple region growing process which starts at the skeleton and proceeds outward (this can be though of as a

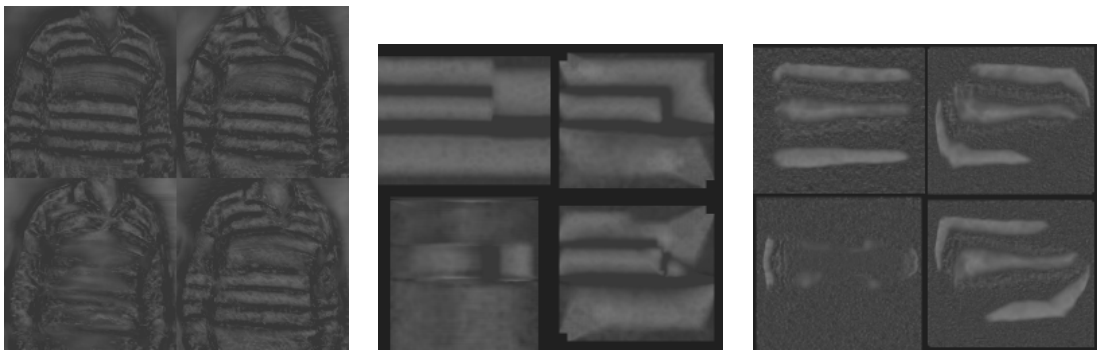


Figure 21: Inertia surfaces for three images at four orientations (clockwise 12, 1:30, 3 and 4:30). Note that *exactly* the same lisp code (without changing the parameters) was used for all the images. *From Left to Right:* Shirt image, Ribbon image, Blob image.

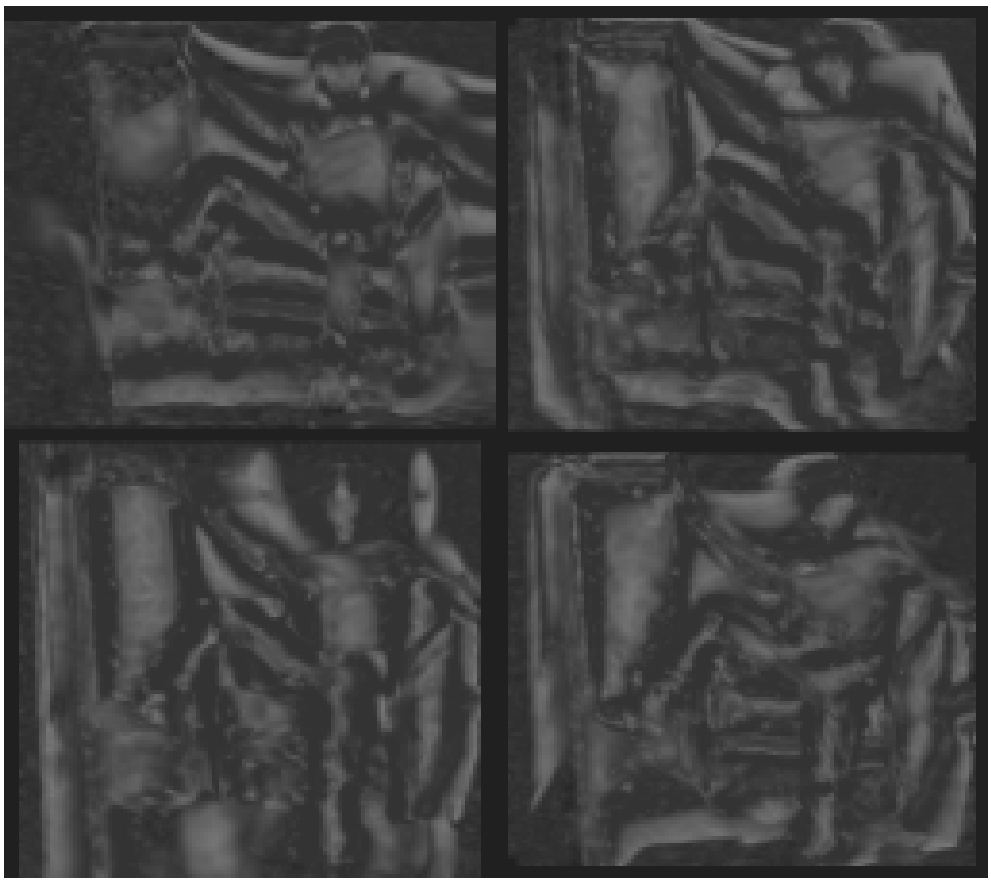


Figure 22: Inertia surfaces for the person image at four orientations. Note that *exactly* the same lisp code (without changing the parameters) was used for these images and the others shown in this paper.

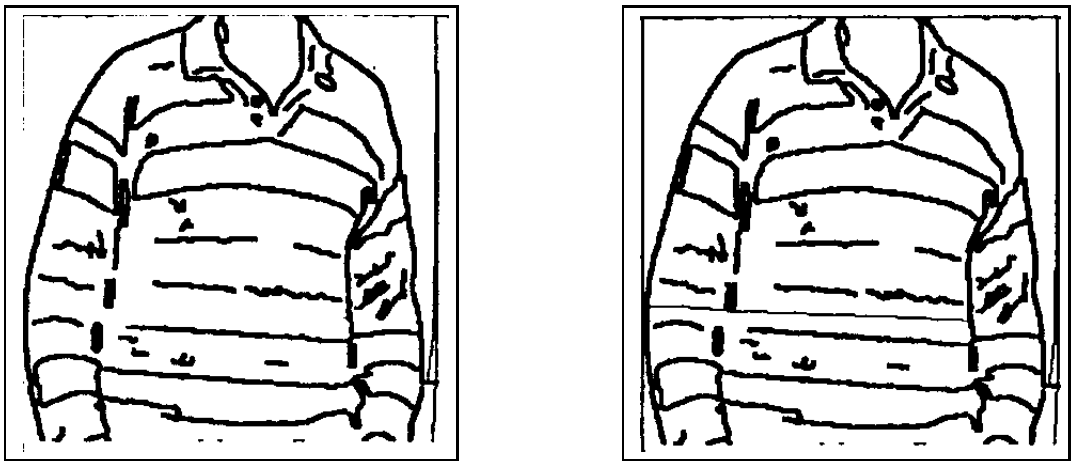


Figure 23: Most salient Curved Inertia Frame obtained in the shirt image. Note that our scheme recovers the structures at the right scale, without the need of changing any parameters. *Left:* Edge map of shirt image without most salient curved inertia frame. *Right:* With most salient curved inertia frame superimposed.

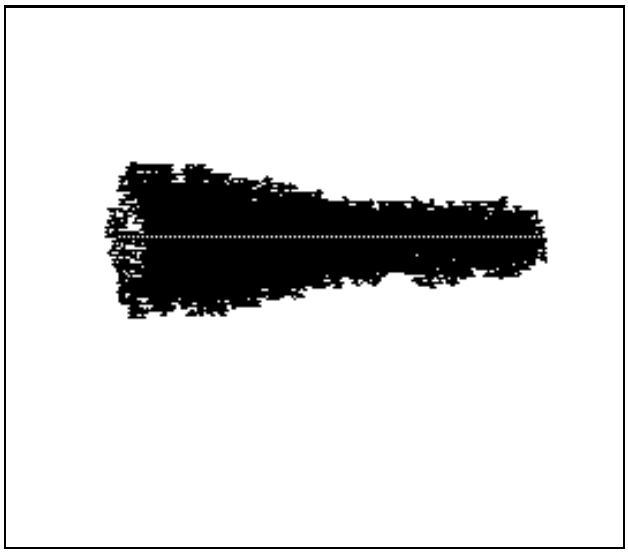


Figure 24: Blob with skeleton obtained using our scheme in the blob image. Note that our scheme recovers the structures at the right scale, without the need of changing any parameters.

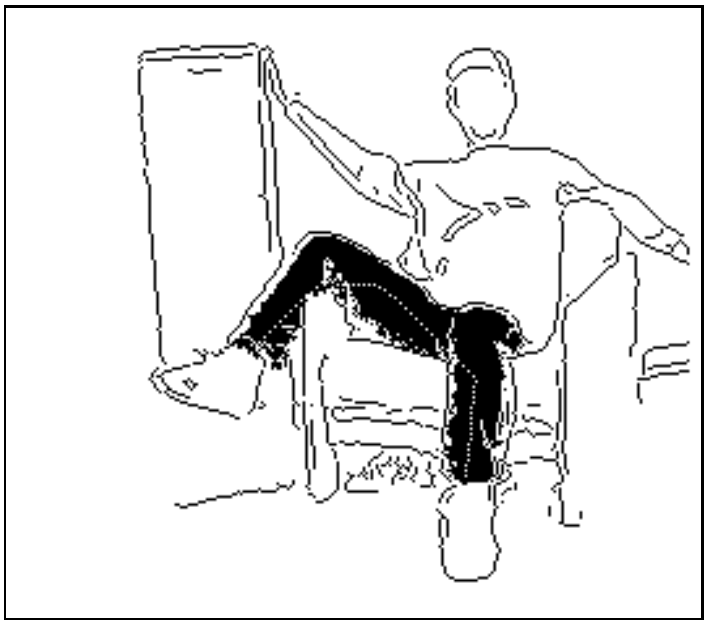


Figure 25: Pants region obtained in person image. The white curve is the Curved Inertia Frames from which the region was recovered.

visual routine [Ullman 1984] operating on the output of the dynamic programming stage or skeleton sketch [Subirana-Vilanova 1990]). This process is very stable because it can use global information provided by the frame such as the average color or the expected size of the enclosing region. See Figures 23, 24, 25, 26, and 27 for some examples of the regions that are obtained. Observe that the shape of the regions is accurate, even at corners and junctions. Note that each region can be seen as an individual test since the computations performed within it are independent of those performed outside it.

12 Discussion: Image brightness is necessary

We have implemented our scheme for color on the Connection Machine. The scheme can be extended naturally to brightness and texture (using the now popular filter-based approaches applied to the image, see [Knuttson and Granlund 1983], [Turner 1986], [Fogel and Sagi 1989], [Malik and Perona 1989], [Bovik, Clark and Geisler 1990], [Thau 1990]). The more cues a system uses, the more robust it will be. In fact, image brightness is crucial in some situations because luminance boundaries do not always come together with color boundaries (e.g. cast shadows).

But, should these different schemes be applied independently? Consider a situation in which a surface is defined by an iso-luminant color edge on one side and by a brightness

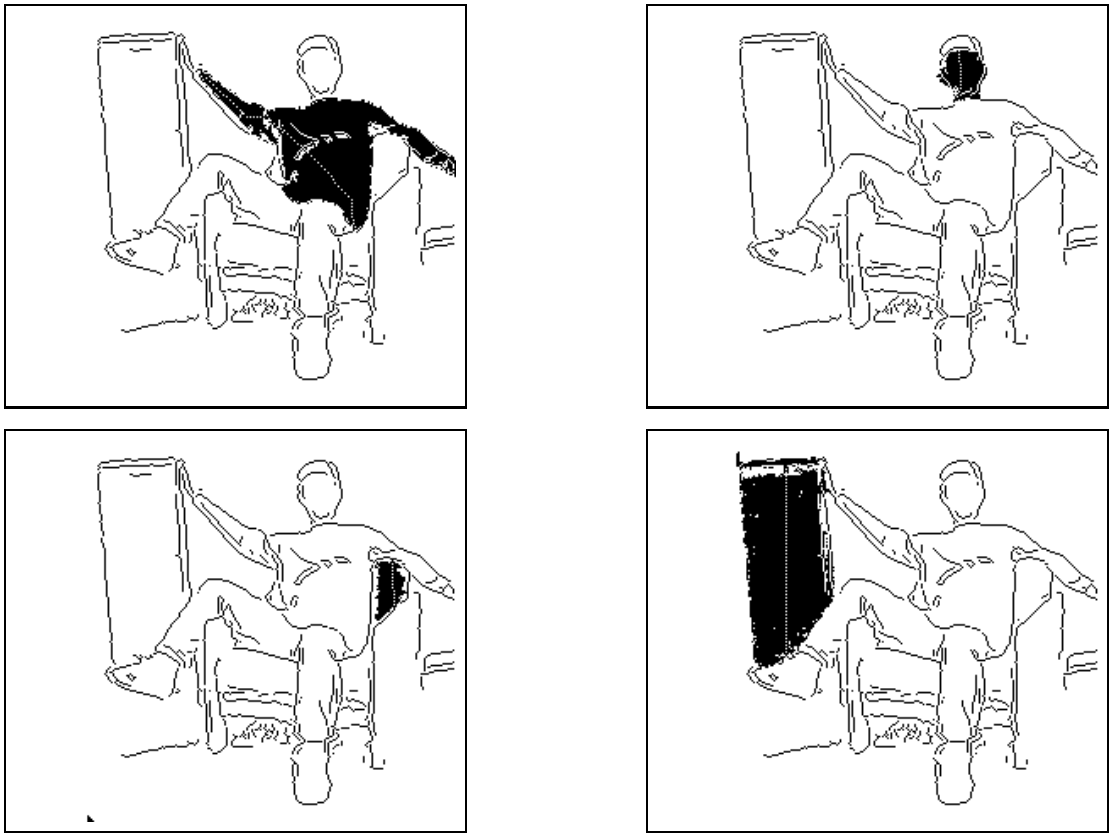
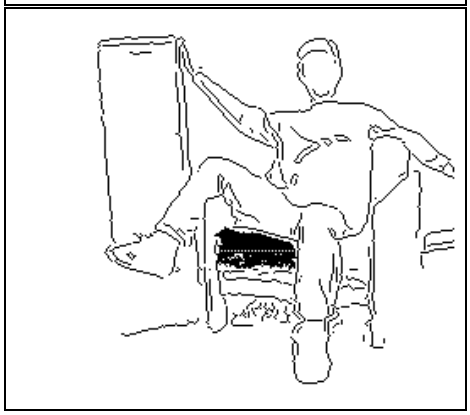


Figure 26: Four regions obtained for the person image. The white curves are the Curved Inertia Frames from which the regions were recovered.

edge (which is not a color edge) on the other. Our scheme would not recover this surface because the two sides of our filter would fail (on one side for the brightness module and on the other for the iso-luminant one). We believe that a combined filter should be used to obtain the inertia values and the tolerated length in this case. The second stage would then be applied only to one set of values. Instead of having a filter with two sides, our new combined filter should have four sides. Two responses on each side, one for color $R_{c,i}$ and one for brightness $R_{b,i}$, the combined response would then be $min(max(R_{b,left}, R_{c,left}), max(R_{b,right}, R_{c,right}))$.



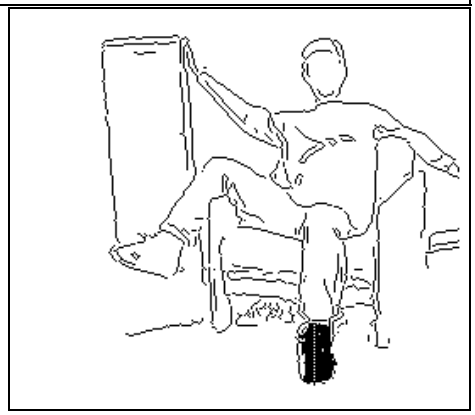


Figure 27: Four other regions obtained for the person image. The white curves are the Curved Inertia Frames from which the regions where recovered.

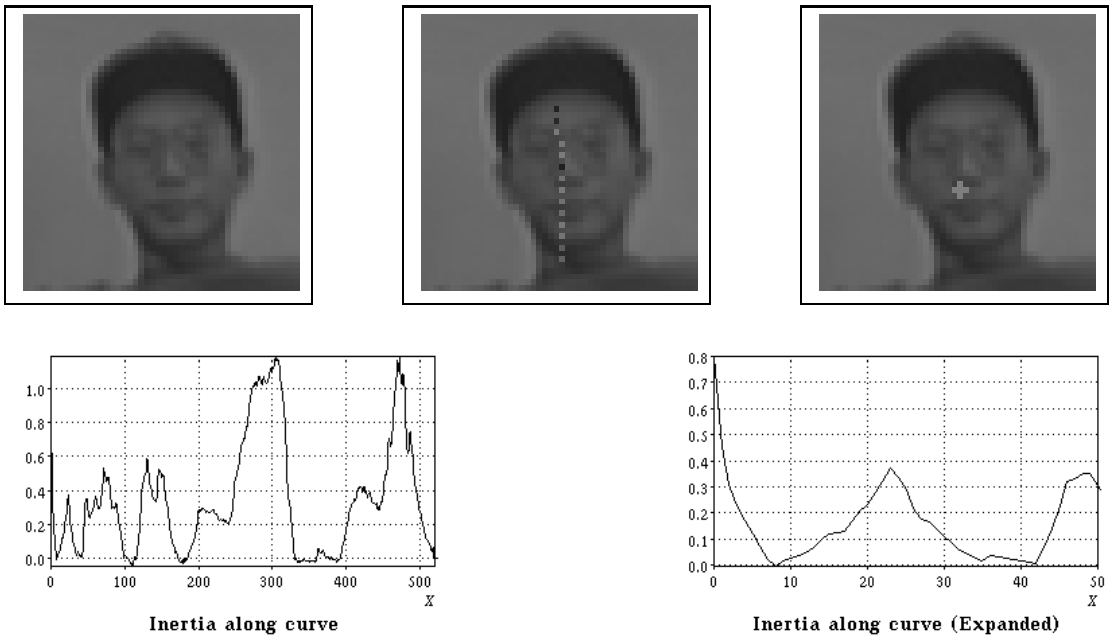


Figure 28: This Figure illustrates how the scheme can be used to guide attention. *Top left:* Close up image of face. *Top center:* Skeletal curve through face. *Top right:* Maximum inertia point on face derived as center of mass of skeletal curve. *Bottom left:* Inertia map along entire skeletal curve, extending beyond the bottom of this image. *Bottom right:* Expanded inertia map focusing on area around face.

Figure 29: Large shapes occlude small ones. From [Kanizsa 1979].

13 What Occludes What?

Our scheme solves the problem of finding different regions by looking at the large structures one by one. The larger structures are the first ones in being recovered, this cuts small structures that are covered by larger structures into different parts. This embodies the constraint that larger structures tend to be perceived as occluding surfaces [Petter 1956]. (See Figure 29).

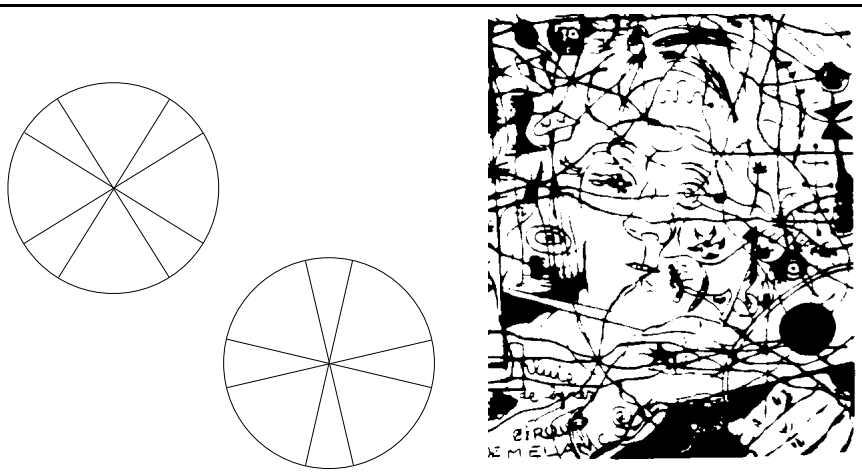


Figure 30: Small structures, whether edges or regions are sometimes more salient. *Left:* From [Rock 1984]. *Right:* Drawing of Miró.

14 Small Is Beautiful Too

As mentioned in [Subirana-Vilanova 1991], the emphasis of our scheme is towards finding large structures. However, this may be misleading as evidenced by Figure 30 where the interesting structure is not composed by individual elements that pop-out in the background. Instead, in this case, what seems to capture our attention can be described as "what is not large". That is, looking for the large structures and finding what is left would recover the interesting structure as if we where getting rid of the background. It is unclear though, if this observation would hold in general. Future research is necessary.

15 Are Edges Necessary?

A central point in this paper has been that the computation of discontinuities should not precede perceptual organization. Further evidence for the importance of perceptual organization is provided by an astonishing result obtained recently by [Cumming, Hurlbert, Johnson and Parker 1991]: when a textured cycle of a sine wave in depth (the upper half convex, the lower half concave) is seen rotating both halfs may appear convex[6], despite the fact that this challenges rigidity[7] (in fact, a narrow band between the two ribbons is seen as moving non-rigidly!). This, at first, seems to violate the rigidity assumption.

[6]The surface can be described by the equation $Z = sin(y)$ where Z is the depth from the fixation plane. The rotation is along the Y-axis by $+/-$ 10 degrees at 1 Hz.

[7]This observation is relevant because it supports the notion that perceptual organization is computed in the image before structure from motion is recovered.

However, these results provide evidence that before finding the structure from motion, the human visual system may segment the image into different components. Within each of this, rigidity can prevail.

Evidence against any form of grouping prior to stereo is provided by the fact that we can understand random dot stereo diagrams even though there is no evidence at all for perceptual groups in one single image. However, it is unclear from current psychological data if this displays take longer time. If they do, one possible explanation (which is consistent with our suggestions) may be that they impair perceptual organization on the individual images and therefore stereo computations. We believe that the effect of such demonstrations has been to focus the attention on stereo without grouping. But perhaps grouping is central to stereo and R.D.S. are just an example of the stability of our stereo system.

A second central point of this paper is that edge detection may not precede perceptual organization. However, there are a number of situations in which edges are clearly necessary as when you have a line drawing image[8] or for the Kanizsa figures. Nevertheless some sort of region processing must be involved also since surfaces are also perceived. We (like others) believe that region-based representations should be sought even in this case. In fact, as we noted in section 2, line drawings are harder to recognize (just like R.D.S. seem to be - but see [Biederman 1988]). The role of discontinuities versus such of regions is still unclear.

16 What's New

In this paper we have argued that early visual processing should seek representations that make regions explicit, not just edges. Furthermore, we have argued that region representations should be computed directly on the image (i.e. not directly from discontinuities). These suggestions can be taken further to imply that an attentional "coordinate" frame (which corresponds to one of the perceptual groups obtained) is imposed in the image prior to constructing a description for recognition (see also [Subirana-Vilanova and Richards 1991]). We have provided some motivation by listing both, a number of problems with alternatives approaches and arguments in favor of region-based schemes.

Our scheme suggests that vision may start by computing a set of features all over the image (corresponding to the inertia values and the tolerated length). This can be thought of as "smart" convolutions of the image with suitable filters plus some simple non-linear processing. In fact, recently filter-based approaches to texture have been presented [Knuttson and Granlund 1983], [Turner 1986], [Fogel and Sagi 1989], [Malik and Perona 1989], [Bovik, Clark and Geisler 1990], stereo [Kass 1983], [Jones and Malik 1990] brightness

[8] Although note that each line has 2 edges (not just one), generally it is assumed that when we look at such drawings we ignore one of the edges. An alternative possibility is that our visual system assembles a region-based description from the edges without merging them.

edge detection [Canny 1986], [Morrone, Owens and Burr 1987, 1990], [Freeman and Adelson 1990] and motion [Heeger 1988]. (See also [Abramatic and Faugeras 1982], [Marrone and Owens 1987]). Our proposal differs from theirs in the non-linear filter proposed and in the use of the filter output to look for ridges and regions, not discontinuities.

This has been the motivation for designing a new non-linear filter for ridge-detection. Our ridge detector has a number of advantages over previous ones since it selects the appropriate scale at each point in the image, does not respond to edges, can be used with brightness as well as color data, is tolerant to noise and can handle narrow valleys and multiple steps.

The resulting scheme can segment an image without making explicit use of discontinuities and is computationally efficient on the Connection Machine (takes time proportional to the size of the image). The performance of the scheme can in principle be attributed to a number of intervening factors; but we believe that one of the critical aspects of the scheme (and one of the contributions of this paper) is our ridge-detector. Running the scheme on the edges or using simple gabor filters would not yield comparable results. The effective use of color makes the scheme very robust but we believe that comparable results would be obtained on brightness or texture data.

References

[1] J-F. Abramatic and O. Faugeras. Sequential convolution techniques for image filtering. *IEEE Trans. Acoust., Speech, Signal Processing*, 30(1):1–10, 1982.

[2] N. Badler and R. Bajcsy. Three-dimensional representations for computer graphics and computer vision. *Computer Graphics*, 12:153–160, 1978.

[3] D. Beymer. Finding junctions using the image gradient. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, pages 720–721, Ann Arbor, MI, 1991.

[4] I. Biederman and G. Ju. Surface versus edge-based determinants of visual recognition. *Cognitive Psychology*, 20:38–64, 1988.

[5] T.O. Binford. Visual perception by a computer. In *Proceedings of the IEEE Conf. on Systems and Controls, Miami*, 1971.

[6] Bovik, Clark, and Geisler. Multichannel texture analysis using localized spatial filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-12:56–65, 1990.

[7] M. Brady. Seeds of perception. In *Proceedings of the Third Alvey Vision Conference, (Cambridge University, 15–17 September)*, pages 259–266. The University of Sheffield Printing Unit, 1987.

[8] R.A. Brooks. Symbolic reasoning among 3-D models and 2-D images. *Artificial Intelligence*, 17:285–348, 1981.

[9] R.A. Brooks, G. Russell, and T. Binford. The acronym model based vision system. In *Proceedings IJCAI*, pages 105–113, 1979.

[10] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8:679–698, 1986.

[11] P. Cavanagh. Reconstructing the third dimension: Interactions between color, texture, motion, binocular disparity, and shape. *Computer Vision, Graphics, and Image Processing*, 37:171–195, 1987.

[12] J.J. Clark. Singularity theory and phantom edges in scale space. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-10:720–727, 1988.

[13] D. Clemens. *Region-based feature interpretation for recognizing 3D models in 2D images.* PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 1991.

[14] B.G. Cumming, A.C. Hurlbert, E.B. Johnson, and A.J. Parker. Effects of texture and shading on the kde. In *The Association for Research in Vision and Ophthalmology. Annual Meeting Abstract Issue. Vol 32, NO. 4*, page 1277, Bethesda, Maryland 20814-3928, 1991.

[15] I. Fogel and D. Sagi. Gabor filters as texture discriminators. *Biological Cybernetics*, 61:103–113, 1989.

[16] D. Forsyth and A. Zisserman. Mutual illumination. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, pages 466–473, 1989.

[17] W.T. Freeman and E. H. Adelson. Steerable filters for early vision, image analysis and wavelet decomposition. In *Third International Conference on Computer Vision*, pages 406–415. IEEE Computer Society, 1990.

[18] D. Geiger and T. Poggio. An optimal scale for edge detection. In *Proceedings IJCAI*, pages 745–748, Milan, Italy, 1987.

[19] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6:721–741, 1984.

[20] M.A. Gennert. Detecting half-edges and vertices in images. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, pages 552–557, 1986.

[21] G. Giraudon and R. Deriche. On corner and vertex detection. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, pages 650–655, Lahaina, Maui, Hawaii, 1991.

[22] W. Eric L. Grimson. *From Images to Surfaces.* MIT Press, Cambridge, Mass., 1981.

[23] W.E.L. Grimson. *Object Recognition By Computer: The Role Of Geometric Constraints*. The MIT Press, Cambridge and London, 1990.

[24] A. Guzman. Decomposition of a visual scene into three-dimensional bodies. In *Proc. AFIPS 1968 Fall Joint Computer Conference*, pages 291–304, 1968.

[25] A. R. Hanson and E. M. Riseman. Segmentation of natural scenes. *CVS*.

[26] R.M. Haralick and L.G. Shapiro. Image segmentation techniques. *Computer Vision, Graphics, and Image Processing*, 29:100–132, 1985.

[27] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of the Fourth Alvey Vision Conference (Manchest er University, 31st August–2nd September)*, pages 147–152. The University of Sheffield Printing Unit, 1988.

[28] D. Heeger. Optical flow from spatiotemporal filters. In *Proceedings of the First International Conference on Computer Vision*, pages 181–190, 1988.

[29] H. Helson. Fundamental principles in color vision. *Journal of Exp. Psychol.*, 23:439–471, 1938.

[30] B. K. P. Horn. Understanding image intensities. *Artificial Intelligence*, 8, 1977.

[31] S. L. Horowitz and T. Pavlidis. Picture segmentation by a direct split and merge procedure. pages 424–433, August 1974.

[32] D.P. Huttenlocher and P. Wayner. Finding convex edge groupings in an image. TR 90-1116, Department of Computer Sciences, Cornell University, Ithaca, New York, 1990.

[33] D.W. Jacobs. Grouping for recognition. A.I. Technical Report No. 1117, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, 1989.

[34] A. Jepson and W. Richards. Integrating vision modules. *IEEE Systems and Cybernetics*, 1991. Special Issue on Assimilation. Editor: A. Jain. To appear.

[35] D. Jones and J. Malik. Computational stereopsis–beyond zero-crossings. *Invest. Ophtalmol. Vis. Sci. (Supplement)*, 31(4):529, 1990.

[36] D. B. Judd and G. Wyszecki. *Color in Business, Science and Industry (Third Edition)*. Wiley, New York, 1975.

[37] D.B. Judd. Appraisal of land's work on two-primary color projections. *Journal of the Optical Society of America*, 50:254–268, 1940.

[38] G. Kanizsa. *Organization in Vision*. Praeger, 1979.

[39] M. Kass. Computing visual correspondence. In *From Pixels to Predicates*, pages 78–92. Ablex Publishing Corporation, Norwood, NH, 1986.

[40] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, pages 321–331, 1988.

[41] H. Knuttson and G.H. Granlund. Texture analysis using two-dimensional quadrature filters. In *Workshop on Computer Architecture for Pattern Analysis and Image Database Management*, pages 206–213, IEE Computer Society, 1983.

[42] J.J. Koenderink. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.

[43] A. Koffka. *The principles of Gestalt psychology.* Harcourt, Brace, New York, 1940.

[44] W. Kohler. *Dynamics in psychology.* Liveright, New York, 1940.

[45] A.F. Korn. Toward a symbolic representation of intensity changes in images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(5):610–625, 1988.

[46] D.G. Lowe. *Perceptual Organization and Visual Recognition.* PhD thesis, Standford University, 1984.

[47] D.G. Lowe. Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence*, 31:355–395, 1987.

[48] Y. Lu and R.C. Jain. Behavior of edges in scale space. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-11:337–356, 1989.

[49] J.V. Mahoney. Image chunking: defining spatial building blocks for scene analysis. A.I. Technical Report No. 980, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, 1987.

[50] J. Malik and Z. Gigus. A model for curvilinear segregation. *Invest. Ophtalmol. Vis. Sci. (Supplement)*, 32(4):715, 1991.

[51] J. Malik and P. Perona. A computational model of texture perception. Report No. UCB-CSD 89-491, Computer science division (EECS), University of California, Berkeley, CA, 1989.

[52] J. Malik and P. Perona. Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America*, 7(5):923–932, 1990.

[53] D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* W.H. Freeman and Company, New York, 1982.

[54] D. Marr and E. Hildreth. Theory of edge detection. *Proceedings of the Royal Society of London*, B(207):187–217, 1980.

[55] J.L. Marroquin. Human perception of structure. Master's thesis, Massachusetts Institute of Technology, 1976.

[56] G. Medioni and Y. Yasumoto. Corner detection and curve representations using cubic b-splines. *Computer Vision, Graphics, and Image Processing*, 39:267–278, 1987.

[57] M.C. Morrone and D.C. Burr. A model of human feature detection based on matched filters. In P. Dario and G. Sandini, editors, *Robots and biological systems*. Academic Press, 1990.

[58] J.A. Noble. Morphological feature detection. In *Proceedings of the First International Conference on Computer Vision*, pages 112–116, Dec 1988.

[59] A. Pentland. Parallel part segmentation for object recognition. Technical Report 108, Vision Sciences, Media Laboratory, Massachusetts Institute of Technology, Cambridge, MA, 1988.

[60] P. Perona and J. Malik. Scale space and edge detection using anisotropic diffusion. In *Proceedings of a Workshop on Computer Vision, Published by IEEE Computer Society Press, Washington, DC.*, pages 16–22, Miami Beach, FL, 1987.

[61] P. Perona and J. Malik. Detecting and localizing edges composed of steps, peaks and roofs. In IEEE Computer Society, editor, *Third International Conference on Computer Vision*, pages 52–57, 1990.

[62] G. Petter. Nuove ricerche sperimentali sulla totalizzazione percepttiva. *Rivista di psicologia*, 50:213–227, 1956.

[63] S.M. Pizer, C.A. Burbeck, and J.M. Coggins. Object shape before boundary shape: scale-space medial axes. In Springer-Verlag, editor, *Proceedings of a NATO Workshop Shape in Picture*, 1992.

[64] T. Poggio, E.B. Gamble, and J. Little. Parallel integration of vision modules. *Science*, 242:436–440, 1988.

[65] J. Ponce and M. Brady. Towards a surface primal sketch. A.I. Memo No. 824, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1985.

[66] I. Rock. *Perception*. Sci. American Library, 1984.

[67] J.M. Rubin and W.A. Richards. Colour vision and image intensities: When are changes material. Technical Report AIM–631, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1981.

[68] T. Ryanm and C. Schwartz. Speed of perception as a function of mode of representation. *American Journal of Psychology*, 69:60–69, 1956.

[69] T. Sanger. Stereo disparity computation using gabor filters. *Biological Cybernetics*, 59:405–418, 1988.

[70] B.G. Schunck. Edge detection with gaussian filters at multiple scales. In *Proceedings of a Workshop on Computer Vision, Published by IEEE Computer Society Press, Washington, DC.*, pages 208–210, Miami Beach, FL, 1987.

[71] A. Sha'ashua and S. Ullman. Structural saliency: The detection of globally salient structures using a locally connected network. In *Proceedings of the First International Conference on Computer Vision*, pages 321–327, 1988.

[72] A. Singh and M. Shneier. Grey level corner detection: A generalisation and a robust real time implementation. *Computer Vision, Graphics and Image Processing*, 51:54–69, 1990.

[73] J.G. Snodgrass and M. Vanderwart. A standarized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Perception and Performance*, 6(2):174–215, 1980.

[74] J.B. Subirana-Vilanova. Curved inertia frames and the skeleton sketch: finding salient frames of reference. In *Proceedings of the First International Conference on Computer Vision*, pages 702–708. IEEE Computer Society Press, 1990.

[75] J.B. Subirana-Vilanova. The skeleton sketch: finding salient frames of reference. In *Proceedings Image Understanding Workshop*, pages 399–414. Morgan and Kaufman, 1990.

[76] J.B. Subirana-Vilanova. Curved inertia frames: Perceptual organization and attention using convexity and symmetry. A.I. Memo No. 1137, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1991.

[77] J.B. Subirana-Vilanova. On contour texture. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, pages 753–754, Ann Arbor, MI, 1991.

[78] J.B. Subirana-Vilanova and W. Richards. Figure-ground in visual perception. In *The Association for Research in Vision and Ophthalmology. Annual Meeting Abstract Issue. Vol 32, NO. 4*, page 697, Bethesda, Maryland 20814-3928, 1991.

[79] K. Sung. A vector signal processing approach to color. AI-TR No. 1349, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1992.

[80] D. Terzopoulos. Regularization of inverse visual problems involving discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(4), July 1986.

[81] R.S. Thau. Illuminant precompensation for texture discrimination using filters. In *Proceedings Image Understanding Workshop*, pages 179–184, Pittsburgh, Pennsylvania, 1990. Morgan Kaufman Publishers Inc., San Mateo, CA.

[82] M. Turner. Texture discrimination by gabor functions. *Biological Cybernetics*, 55:71–82, 1986.

[83] S. Ullman. Filling in the gaps: The shape of subjective contours and a model for their generation. *Biological Cybernetics*, 25:1–6, 1976.

[84] S. Ullman. Visual routines. *Cognition*, 18, 1984.

[85] D.L. Waltz. Understanding line drawings of scenes with shadows. In P. Winston, editor, *The Psychology of Computer Vision*. McGraw-Hill, New York, 1972.

[86] M. Wertheimer. Principles of perceptual organization. In B. Beardslee and M. Wertheimer, editors, *Readings in perception*. Van Nostrand, Princeton, 1958. Originally published in 1923.

[87] A.P. Witkin. Scale-space filtering. In *Proceedings IJCAI*, pages 1019–1022, 1983.

[88] A.P. Witkin. Scale space filtering: a new approach to multi-scale description. In *Proceedings Image Understanding Workshop*, pages 79–95, Pittsburgh, Pennsylvania, 1984. Morgan Kaufman Publishers Inc., San Mateo, CA.

[89] A.P. Witkin and J.M. Tenenbaum. On the role of structure in vision. In J. Beck, B. Hope, and A. Rosenfeld, editors, *Human and Machine Vision*. Academic Press, New York, 1983.

[90] S. Zhong and S. Mallat. Compact image representation from multiscale edges. In IEEE Computer Society, editor, *Third International Conference on Computer Vision*, pages 406–415, 1990.

[91] X. Zhuang, T.S. Huang, and H.H. Chen. Multi-scale edge detector using gaussian filtering. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, pages 558–563, 1986.

[92] S.W. Zucker, A. Dobbins, and L. Iverson. Two stages of curve detection suggest two styles of visual computation. *Neural Computation*, 1(1):68–81, 1989.