Ridge-Detection for the Perceptual Organization Without Edges

J. Brian Subirana-Vilanova and Kah Kay Sung

M.I.T. Artificial Intelligence Laboratory
Cambridge, MA, 02139

Abstract

We present a novel vector ridge detector which is designed to automaticaly 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.

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 on 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 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 [Haralick and Shapiro 1985]. These schemes proceed by laying a few "seeds" in the image and then "growing" these until a complete region is found. The growing

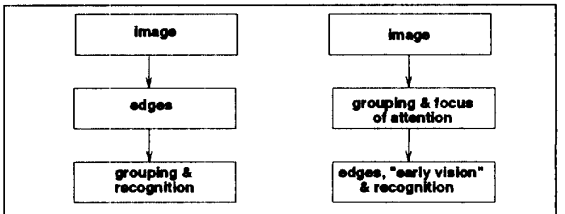

Figure 1: Two different views on the role of perceptual organization.

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], [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, 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

0-8186-3870-2/93 $3.00 © 1993 IEEE

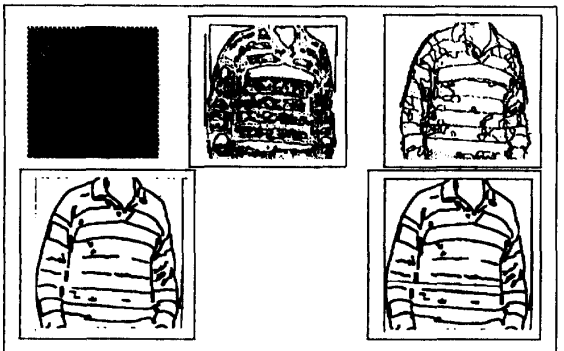

Figure 2: *Top left:* An image of a shirt. *Top center:* Original seeds for a region growing segmentation algorithm. *Top right:* Final segmentation obtained using a region growing algorithm. *Below:* Edge image and skeleton obtained on color image with C.I.F.. The skeleton can be used to find the shirt ribbon by growing a region starting on the skeleton and using parameters based on the whole skeleton.

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

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 which makes 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 and then present our approach. Some results of a version of our scheme implemented on the Connection Machine will be shown at the end of the paper.

2 In Favor of Regions

What is an edge? Unfortunately there is no agreed definition of it. It can be defined in several related ways: as a discontinuity in a certain property, as "something" that looks like a step edge [Canny 1986] (see Figure 3), or by an algorithm (e.g. zero-crossings [Marr and Hildreth 1980]). Characterizing edges has

Figure 4: Edges computed at three different scales for an image of a person. Note that the results are notably different. Which scale is best?

proven to be difficult especially near corners, junctions and when there are edges at multiple scales, noise, or transparent surfaces.

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 that share a common property. In this context, an edge is the border of a region. But 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). But this suffers from problems similar to the ones mentioned for edges.

Thus, regions and edges are two concepts closely related. It is unclear how we should represent the information contained in an image. As regions? As edges? Furthermore, independently of our choice, which structures should we try to recover first? Edges or regions? We believe that computer vision has over emphasized the early computation of discontinuities (whether brightness, texture or color discontinuities). Here are some reasons why exploring the computation of regions (without edges) may be a promising approach (see [Subirana-Vilanova and Sung 1991] for a more comprehensive discussion including some references):

- There is psychological evidence that humans can recognize images with region information better than line drawings.
- Representations which maintain some region information such as the sign-bit of the zero crossings (instead of just the zero crossings themselves) are useful for perceptual organization.
- The performance of most rigid-object schemes is bounded by the complexity of the feature space used for exploring possible matchs. Additional region groups should reduce such complexity.
- Previous research on recognition has focused on rigid objects. Related grouping research has focussed on finding small sets of features with high likelihood of coming from the same object. For non-rigid objects this is not sufficient. Instead, it is necessary to group most of the features coming from a single object. We find it hard to believe that edge-features will be sufficient for bottom-up grouping in this case.
- Scale and stability are recognized as important problems. However, is it the stability and scale

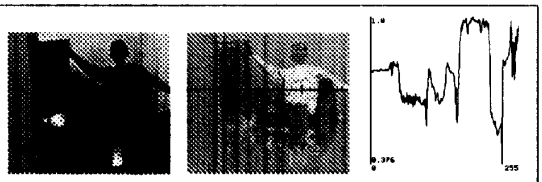

Figure 5: The similarity measure described in Equation 1 is illustrated here for an image of a person. *Left:* Image. *Center:* Similarity measure using as reference color the color of the pixel located at the intersection of the two segments shown. *Right:* Plot of the similarity measure along the long segment using the same reference color.

of an edge? or that of a region that we are interested in? Our scheme addresses stability in terms of objects (not edges). In addition, our scheme commits to one scale corresponding to 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. In fact, color similarity measures are based on vector values and cannot be mapped onto a one-dimensional measure. This makes color perception different from brightness from a computational point of view since not all the one-dimensional techniques used in brightness images extend naturally to higher dimensions. The extension to texture and brigthness is possible since both can be casted naturaly into a two-dimensional vector field.

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 5. In fact, most other measures proposed in the literature can be incorporated in our scheme. 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 $S_{\otimes}$ between pairs of pixels $\vec{c}$ and $\vec{c_R}$. The difference that we used is defined in equation 1 and was taken from [Subirana-Vilanova and Sung 91] ($\otimes$ denotes the *vector cross product* operation) and responds very sensitively to color differences between similar colors.

$$S_{\otimes}(\vec{c}) = 1 - \frac{|\vec{c} \otimes \vec{c_R}|}{|\vec{c}||\vec{c_R}|} \qquad (1)$$

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", $\vec{c_R}$, and a minimum value of 0 to colors that are orthogonal to $\vec{c_R}$ in the RGB vector space. The discriminability of this measure can be seen intuitively by looking at the normalized image in Figure 5.

Figure 6: *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.

4 Problems in Finding Ridges

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.

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).

The hat model has a number of problems which illustrate some useful features of a ridge operator:

- *Scale:* Ridges should be detected at multiple scales.
- *Non-edgeness:* The filter should give no response for a step edge. This property is violated by [Canny 1986].
- *Multiple steps:* 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. However, a ridge detector should also detect small steps which are not well described by a hat. 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 no just scalar values (see Figure 6).
- *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.

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	$F_s\sigma$	Size of Side Filter Lobe.
σ_c	$F_c\sigma$	Reference Color Neighborhood
$\mathbf{c}(x)$	$[\mathcal{R}(x)\,\mathcal{G}(x)\,\mathcal{B}(x)]^t$	Color vector at location x.
$\mathbf{c}_n(x)$	$\mathbf{c}(x)/\lvert\mathbf{c}(x)\rvert$	Normalized Color at x.
$\mathbf{c}_r(x)$	$\int_{-\sigma_c}^{\sigma_c} \frac{1}{\sqrt{2\pi}\sigma_c} e^{-\frac{r^2}{2\sigma_c^2}} \mathbf{c}_n(x+r)\,dr$	Reference Color at x
$\mathcal{F}_L(r)$	$\frac{r+\sigma}{\sigma^3\sqrt{2\pi}} e^{-\frac{(r+\sigma)^2}{2\sigma^2}}$ $-\sigma < r < \sigma$ $\frac{r+\sigma}{\sigma_s^3\sqrt{2\pi}} e^{-\frac{(r+\sigma)^2}{2\sigma_s^2}}$ $-(\sigma+\sigma_s) < r < -\sigma$ 0 otherwise	Left Half of Filter
$\mathcal{F}_R(r)$	$\mathcal{F}_L(-r)$	Right Half of Filter
$\mathcal{I}_L(x)$	$\int_{-(\sigma+\sigma_s)}^{\sigma} S_\Theta(\mathbf{c}_r(x), \mathbf{c}_n(x+r))\mathcal{F}_L(r)\,dr$	Inertia from Left Half
$\mathcal{I}_R(x)$	$\int_{-\sigma}^{\sigma+\sigma_s} S_\Theta(\mathbf{c}_r(x), \mathbf{c}_n(x+r))\mathcal{F}_R(r)\,dr$	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.

- *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.

Previous schemes to find ridges often fail to consider one or more of the above issues. For example, [Canny 1986] and [Subirana-Vilanova 1990] assumed that ridges appear at just one scale. While this is a good working assumption for edges, it is simply not true in most images (except in simple ones like text images where objects have a fixed width).

5 A Color/Vector 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

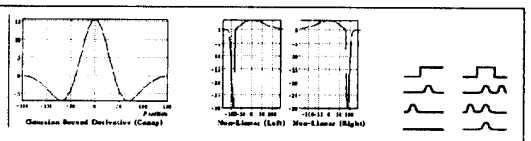

Figure 7: *Left:* Gaussian second derivative, an approximation to Canny's optimal ridge detector. *Center:* Individual one-dimensional masks used by our operator. *Right:* Response for an edge (left) and a ridge (right) for: the left operator of our ridge detector (second row), the right operator (third row) and the combined minimum of the two responses (fourth row). In all cases, sigma is roughly the size of the ridge. This is a qualitative demonstration. It illustrates why combining the responses of two filters yields a positive response for a ridge but no response for an edge.

to edges, so will the combined filter do - see Figure 10).

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* of the two responses. This is the basic idea behind our approach (see Figure 7). Figure 10 illustrates 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 1 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].

Our approach uses two filters (see profile in Figure 7), 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 is tricky because there is no clear notion of what is positive and what is negative in vector quantities nor in steps. 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. Scalar deviations from this reference color are computed as

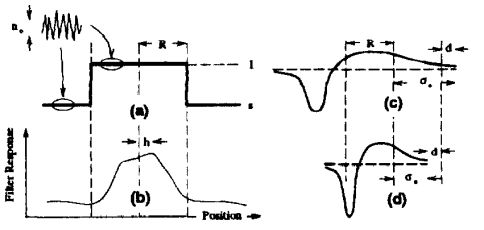

Figure 8: Notations for (a): Ridge profile, (b): Spatial Localization, and (c) and (d): Scale localization.

defined in section 3.

6 Filter Characteristics

This section summarizes the main results of our filter's *optimum scale response*, *scale localization* and *spatial localization* characteristics under varying degrees of noise. We refer the interested reader to [Subirana-Vilanova and Sung 91] for details. Scale localization measures the closeness in value between the actual width of a ridge and the filter's optimum mask size at the ridge center. Spatial localization measures proximity between the filter's peak response location and the actual ridge center. We shall see that both these measures remain remarkably stable even at noticeably high noise levels.

For simplicity, we use scalar ridge profiles instead of color ridge profiles. Our filter notations are similar to those given in Table 1. In particular, σ denotes the main lobe's width (or scale) and F_s denotes the filter's main lobe to side lobe width ratio. The following are some additional notations (see Figure 8): R is the width of the ridge we are detecting, $(1-s)$ is the ridge height, $d = |R - \sigma_o|$ is the size difference between the ridge width and the optimal filter scale at the ridge center (for scale localization), and h is the distance between the actual ridge center and peak location of the filter's *all scales* ridge response (for spatial localization). Noise is *white* with variance n_o^2.

6.1 Response and Localization

Let x be distance from the ridge center and $\sigma_o(x)$ be the maximum response filter scale at x. It can be shown that the *all scales optimum filter response* for a noiseless ridge is:

$$\texttt{Opt}(x,R) = \frac{1}{\sqrt{2\pi}}[(F_s s-1)(e^{-2}-1) - (1-s)(1-e^{-\frac{(\sigma_o(x)+x-R)^2}{2\sigma_o(x)^2}})], \quad (2)$$

where $\sigma_o(x) \approx (R+x)$ falls within the following bounds:

$$\frac{R+x}{1+\frac{\sqrt{2}}{F_s}\ln(\frac{F_s}{F_s-1-e^{-2}})} < \sigma_o < (R+x). \quad (3)$$

Fixing $\sigma_o(x)$ as $(R+x)$ and differentiating Equation 2 with respect to x, we see that $\texttt{Opt}(x,R)$ indeed peaks at $x=0$.

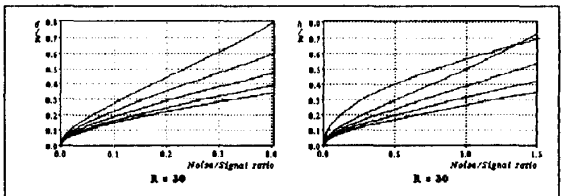

Figure 9: Comparison of **Left:** relative scale error (d/R), and **Right:** relative spatial error (h/R), as a function of *noise to signal ratio* $(n_o/(1-s))$ between our filter and the Canny ridge operator. For the d/R plot, curves from top to bottom are those of: $F_s = 16$, $F_s = 8$, $F_s = 4$, $F_s = 2$, and Canny. For the h/R plot, curves from top to bottom are: Canny, $F_s = 16$, $F_s = 8$, $F_s = 4$ and $F_s = 2$. Curves are very similar for other values of R.

For *scale localization*, we want to estimate d/R, the *relative* scale error due to noise, in terms of the *noise to signal ratio* $n_o/(1-s)$. It can be shown that:

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

$$\text{where :} \quad K = -\ln(1 - \frac{n_o}{1-s}\frac{1}{1-e^{-2}}\sqrt{\frac{1+F_s}{8R\sqrt{\pi}}}). \quad (4)$$

Similarly, for *spatial localization*, we want to establish a magnitude bound for h/R in terms of the *noise to signal ratio*. We have:

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

$$\text{where :} \quad K = -\ln(1 - \frac{n_o}{1-s}\sqrt{\frac{1+F_s}{8R\sqrt{\pi}}}). \quad (5)$$

6.2 Comparison with 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 1986] 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:

$$C(r,\sigma) = \frac{1}{\sqrt{2\pi}\sigma^3}(\sigma^2 - r^2)e^{-\frac{r^2}{2\sigma^2}}. \quad (6)$$

For *scale localization*, the Canny ridge operator yields the following equation that implicitly relates d/R to $\frac{n_o}{1-s}$:

$$\frac{n_o}{1-s} = \sqrt{\frac{16R}{3\sqrt{\pi}}}\left[e^{-\frac{1}{2}} - \frac{R}{R+d}e^{-\frac{R^2}{2(R+d)^2}}\right]. \quad (7)$$

For *spatial localization*, we get:

$$\frac{n_o}{1-s} \approx \sqrt{\frac{4R}{3\sqrt{\pi}}}[e^{-\frac{1}{2}} - \frac{1}{\sigma_o}e^{-\frac{R^2+h^2}{2\sigma_o^2}} (R\cosh(\frac{Rh}{\sigma_o^2}) - h\sinh(\frac{Rh}{\sigma_o^2}))], \quad (8)$$

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

We see from Figure 9 that at typical F_s ratios, our filter's scale and spatial localization characteristics are comparable to those of the Canny ridge operator.

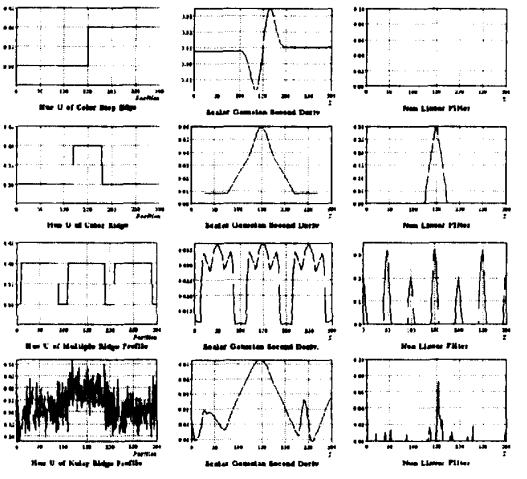

Figure 10: *First column:* Different input signals. *Second column:* Output given by second derivative of the gaussian. *Third column:* Output given by our ridge detector using reference color. The *First row* shows the result of a single scale filter application on an edge profile. Notice that there is no edge response. The *Second, Third and Fourth rows* are results of a multiple scale filter application on three different profiles. Note that no scale parameter is involved in any case.

7 Previous work on perceptual organization without edges

The scheme 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 because it is the only existing scheme that can compute global regions directly on the image without imposing a three-dimensional representation of the data.

[Subirana-Vilanova 1990]'s scheme proceeds in three stages. In the first one, it computes two local measures

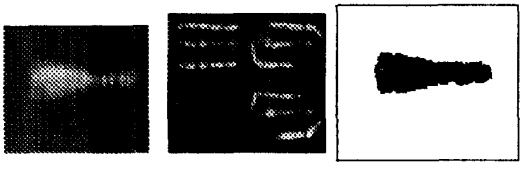

Figure 11: *Left:* Blob image. *Center:* Inertia surfaces for blob image at four orientations (clockwise 12, 1:30, 3 and 4:30). Note that *exactly* the same lisp code (without changing the parameters) was used for all the images. *Right:* Blob with skeleton obtained using our scheme in the bolb image. Note that our scheme recovers the structures at the right scale, without the need of changing any parameters.

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 9, 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. The tolerated length depends also on the curvature of the curve $C(t)$ so that more curvature is permited on skeletons going through narrow regions of the image [Subirana-Vilanova 1991].

$$S_L = \int_0^L \mathcal{I}(l)\rho^{\int_0^l \frac{1}{\alpha\mathcal{T}(t)}dt}dl \qquad (9)$$

In the second stage, the scheme computes the skeleton which yields the maximum saliency using dynamic programming.

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

In the next section we will present some results showing the robustness of the scheme in the presence of noisy shapes.

8 Results

We have tested our scheme (filter + network) extensively, Figure 10 shows that our filter produces sharper and more stable ridge responses than the second derivative of a gaussian filter. First, our filter localizes all the ridges for a single ridge, for multiple or step ridges and for noisy ridges. In contrast, the second derivative of the gaussian fails under the presence of multiple or step ridges. Second, the scale chosen by our operator matches the underlying data closely. And third, our filter does not respond to edges while the second derivative of the gaussian does.

Figure 11 shows the directional output (i.e. inertia surfaces) of our filter on the person image. 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.

The inertia surfaces and the tolerated length are the output of the first stage of our scheme. In the

Figure 12: 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.

second stage we use these to compute the Curved Inertia Frames (see [Subirana-Vilanova 1990]) as shown in Figures 2, 11 and 13. 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 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 2, 11 and 13 for some examples of the regions that are obtained. Observe that the shape of the regions is accurate, even at corners or 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.

9 Discussion: Brightness and Edges Are Necessary

A central motivating point is that edge detection may not precede perceptual organization (see [Subirana-Vilanova and Sung 1991] and [Subirana-Vilanova and Richards 1991] for a more detailed discussion). However, there are a number of situations in which edges are clearly necessary as when you have a line drawing or in illusory contours. Nevertheless some sort of region processing must be involved since surfaces are also perceived. We (like others) believe that region-based representations should be sought even in this case.

We have implemented our scheme for color on the Connection Machine. The scheme can be extended naturally to brightness and texture. 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,

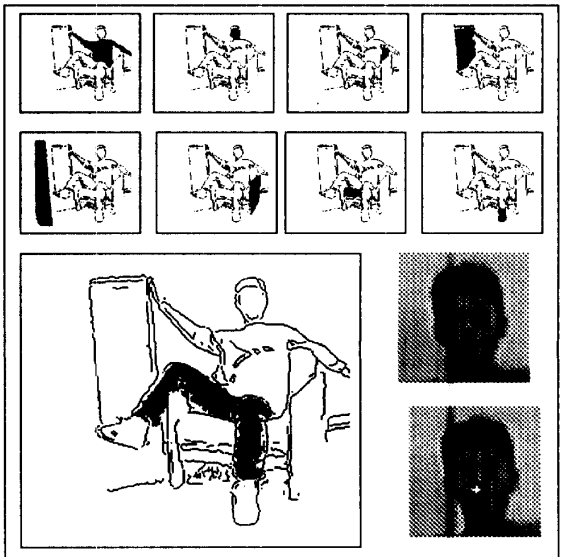

Figure 13: Regions obtained for the person image. The white curve on the bottom image is the Curved Inertia Frame from which the pant region was recovered. **Bottom Right:** An illustration of how the scheme can be used to guide attention. Close-up image of face and maximum intertia point.

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 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). A combined filter could 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 has four sides. Two responses on each side, one for color $R_{c,i}$ and one for brightness $R_{b,i}$, the combined response is then:

$$min(max(R_{b,left}, R_{c,left}), max(R_{b,right}, R_{c,right})).$$

10 Discussion: Large versus Small

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, cutting into small parts the structures that are occluded. This embodies the constraint that larger structures tend to be perceived as occluding surfaces. (See Figure 14.)

The emphasis of our scheme is towards finding large structures. However, this may be misleading as shown in Figure 14 where the interesting structure is not composed by individual elements that pop-out in the background. Instead, what seems to capture our attention can be described as "what is not large". That is, look-

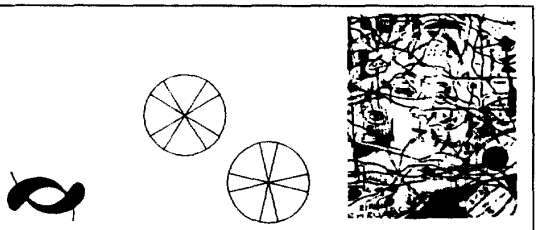

Figure 14: *Left:* Large shapes occlude small ones. From [Kanizsa 1979]. *Center:* Small structures, whether edges or regions are sometimes more salient. From [Rock 1984]. *Right:* Drawing of Miró. This drawing illustrates a similar principle as the center Figure but using edges rather than regions.

ing 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.

11 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 using 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, motion, and stereo have been presented (see references in [Subirana-Vilanova and Sung 1991]). 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[1] and can handle narrow valleys and multiple steps.

[1]We have verified this empirically but we have not performed a detailed study of the effect of noise in our filter. An improved version of the filter may be obtained from such exercise.

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] A. Blake and A. Zisserman. *Visual Reconstruction.* The MIT Press, Cambridge, MA, 1987.

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

[3] 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.

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

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

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

[7] 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.

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

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

[10] J.B. Subirana-Vilanova. Curved inertia frames and the skeleton sketch: finding salient frames of reference. In *Proc. First Intl. Conf. Comp. Vis.*, pages 702–708. IEEE Computer Society Press, 1990.

[11] J.B. Subirana-Vilanova and W. Richards. Perceptual organization, figure-ground, attention and saliency. AIM No. 1218, MIT AT Lab, 1991.

[12] J.B. Subirana-Vilanova and K.K. Sung. Ridge-detection for the perceptual organization without edges. AIM No. 1318, MIT AT Lab, 1991.

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