MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 1137 October, 1991

CURVED INERTIA FRAMES: VISUAL ATTENTION AND PERCEPTUAL ORGANIZATION USING CONVEXITY AND SYMMETRY

J. Brian Subirana-Vilanova

Abstract:
In this paper we present an approach to perceptual organization and attention based on Curved Inertia Frames (C.I.F.), a novel definition of "curved axis of inertia". Such a definition is novel because it is global and can detect curved axes; it can also be used to compute a frame of reference of the shapes in an image useful for non-rigid object recognition or to pull out interesting structures in the image. The scheme assigns a saliency measure to each component of the reference frame that is a measure of its relevance, so that *large, smooth, convex, symmetric and central parts* play a more central role in the description of the shape. One of the remarkable features of the scheme is its tolerance to noisy and spurious data.

Several perceptual phenomena observed in humans such as grouping based on symmetry or convexity and environmental bias in shape description can be supported naturally in this scheme. The scheme also supports other operations such as finding the most "interesting point" or "feature" in the image (for subsequent processing) or defining what is inside and what is outside an object. An extension of the scheme to find long and smooth ridges on an arbitrary surface is presented. The extension is illustrated in the problem of finding salient blobs in images and it is suggested that similar schemes be used in other early and middle level vision tasks.

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 grant S1-801534-2 from Hughes Aircraft Company and by the Advanced Research Projects Agency of the Department of Defense under Army contract DACA76-85-C-0010, and in part by ONR contract N00014-85-K-0124.

1 Introduction

The Problem: Finding Reference Frames

A shape description is an encoding of a shape. A common approach is to describe the points of the shape in a cartesian coordinate reference frame fixed in the image (see Figure 1). An alternative is to center the frame on the shape so that a canonical description can be achieved. For some shapes this can be obtained by orienting the frame of reference along the inertia axis of the shape (see Figure 1). If the objects are elongated and flexible, we suggest another alternative that might be more appropriate, the use of a curved frame of reference (see Figure 2). Recognition can be done using a canonical description of the shape obtained by rotating or "unbending" the shape using the frame as an anchor structure (see Figure 2). For complex shapes, a part decomposition for recognition can be obtained with a skeleton-like frame (e.g. [Connell and Brady 87], see Figure 3). In this paper, we address the problem of finding reference frames (a.k.a. skeletons, symmetry or distance transforms, voronoi diagrams etc.), for a variety of tasks such as recognition, attention, figure-ground and perceptual organization. Our approach is based on Curved Inertia Frames, a novel definition of "curved axis of inertia".

Other Applications Of Reference Frames: Perceptual Organization, Attention, Feature and Corner Detection, Part Segmentation, and Shape Description

The use of reference frames need not be restricted to recognition. Non-recognition examples include: finding an exit path in the maze of Figure 4, finding the corner in Figure 5, finding the blob in Figure 6, determining figure-ground relations in Figure 7 and finding the most interesting object in Figure 17. These examples are closely related to figure-ground relations and perceptual organization (a.k.a. grouping, selection and segmentation), a process that computes regions of the image coming from one single object (of interest if possible), with little detailed knowledge of the particular objects present in the image. The main advantage of our scheme over previously presented perceptual organization schemes [Marroquin 1976], [Witkin and Tenenbaum 1983], [Mahoney 1985], [Harlick and Shapiro 1985], [Lowe 1984, 1987], [Sha'ashua and Ullman 1988], [Jacobs 1989], [Grimson 1990], [Subirana-Vilanova 1990] is that it can find complete curved, symmetric and large structures directly on the edges of the image without requiring features like straight segments or corners. In this context, perceptual organization is related to part segmentation [Hollerbach 1975], [Marr 1977], [Duda and Hart 1973], [Binford 1981], [Hoffman and Richards 1984], [Vaina

and Zlateva 1990], [Badler and Bajcsy 1978], [Binford 1971], [Brooks, Russel and Binford 1979], [Brooks 1981], [Biederman 1985], [Marr and Nishihara 1978], [Marr 1982], [Guzman 1969], [Pentland 1988] and [Waltz 1975] since we are interested in finding an arrangement of structures in the image, not just on finding some of them.

Some Reasons Why Finding Reference Frames Is Not Trivial

Finding reference frames is a straightforward problem for simple geometric shapes such as a square or a rectangle. The problem becomes difficult for shapes that do not have a clear symmetry axis such as a notched rectangle (for some more examples see Figures 2, 8, 9, and 15) and none of the schemes presented previously can handle them successfully. Ultimately, we would like to achieve human-like performance. This is difficult partly because what humans consider to be a good skeleton can be influenced by high-level knowledge (see Figure 8).

Previous Work

The study of reference frames has received considerable attention in the computer vision literature. Reference frames have been used for different purposes (as discussed above) and given different names (e.g. skeletons, voronoi diagrams, symmetry transforms). Previous schemes for computing skeletons fall usually into one of two classes. The first class looks for straight axis, such as the axis of inertia. These methods are global (the axis is determined by all the contour points) and produce a single straight axes. The second class can find a curved axis along the figure, but the computation is based on local information. That is, the axis at a given location is determined by small pieces of contours surrounding this location. Examples of such schemes are, to name but a few, Morphological Filters (see [Serra 82] for an overview), Distance Transforms [Rosenfeld and Pfaltz 68], [Borgefors 86], [Arcelli, Cordella and Levialdi 81], Symmetric Axis Transform [Blum 67], [Blum and Nagel 78] and Smoothed Local Symmetries [Brady and Asada 84], [Connell and Brady 87]. Recently, computations based on physical models have been proposed by [Brady and Scott 88] and [Scott, Turner and Zisserman 89]. In contrast, the novel scheme presented in this paper, which we call Curved Inertia Frames (C.I.F.), can extract curved symmetry axes, and yet use global information.

Outline

The approach that we present for finding skeletons is divided into two successive stages. In Section 3, we present the first stage, in which we obtain two local measures at every point, the *inertia value* and the *tolerated length*, which will provide a local symmetry measure at every point, and for every orientation. This measure is high if locally the point in question appears to be a part of a symmetry axis. This simply means that, at the given orientation, the point is equally distant from two image contours. The symmetry measure therefore produces a map of potential fragments of symmetry curves which we call the inertia surfaces. In Sections 4 and 5, we present the second stage in which we find long and smooth axes going through points of high inertia values and tolerated length. In section 6 we introduce the skeleton sketch and show some results and applications of the scheme, and in section 7 we discuss the relation of our scheme to human perception. We conclude in section 8 by presenting an extension of the scheme to find high, long, and smooth curves on an arbitrary surface. The extension is illustrated on the problem of finding salient blobs in images. In section 8 we also present some limitations of our scheme and a number of topics for future research.

In Appendix I we prove a theorem that shows some strong limitations on the class of measures computable by the computation described in sections 4 and 5.

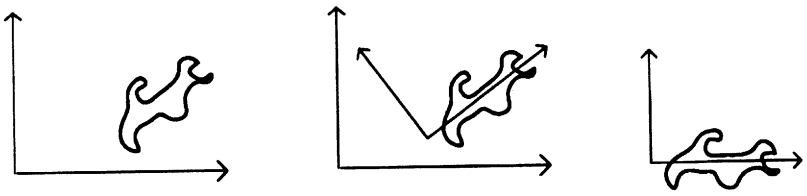

Figure 1: *Left:* a shape described in a image or viewer centered reference frame. *Center:* the same shape with an object centered reference frame superimposed on it. *Right:* a canonical description of the shape.

2 Five Problems With Previous Approaches

Previously presented computations for finding a curved axis generally suffer from one or more of the following problems: first, they produce disconnected skeletons for shapes that deviate from perfect symmetry or that have fragmented boundaries (see

Figure 2: Which two of the three shapes on the left are more similar? One way of answering this question is by "unbending" the shapes using their skeleton as a reference frame, which results in the three shapes on the right. Once the shapes have been unbent, it can be concluded using simple matching procedures that two of them have similar "shapes" and that two others have similar length. We suggest that the recognition of elongated flexible objects can be performed in some cases by transforming the shape to a canonical form and that this transformation can be achieved by unbending the shape using its skeleton as an anchor structure. The unbending presented in this figure was obtained using an implemented lisp program.

Figure 5); second, the obtained skeleton can change drastically for a small change in the shape (e.g. a notched rectangle vs a rectangle) making these schemes unstable; third, they do not assign any measure to the different components of the skeleton that indicates the "relative" relevance of the different components of the shape; fourth, many computations depend on scale, introducing the problem of determining the correct scale; and fifth, it is unclear what to do with curved or somewhat-circular shapes because they do not have a clear symmetry axis.

Consider for example, the Symmetric Axis Transform [Blum 67]. The SAT of a shape is the set of points such that there is a circle centered at the point that is tangent to the contour of the shape at two points and that it does not contain any portion of the boundary of the shape, see [Blum 67] for details. An elegant way of computing the SAT is by using the brushfire algorithm which can be thought of as follows: A fire is lit at the contour of the shape and propagated towards the inside of the shape. The SAT will be the set of points where two fronts of fire meet. The Smoothed Local Symmetries [Brady and Asada 84] are defined in a similar way but, instead of taking the center point of the circle, the point that lies at the center of the segment between the two tangent points is the one that belongs to the SLS and the circle needs not be inside the shape. In order to compute the SAT or SLS of a shape we need to know the tangent along the contours of the shape. Since the tangent is a scale dependent measure so is the SLS. One of the most common problems

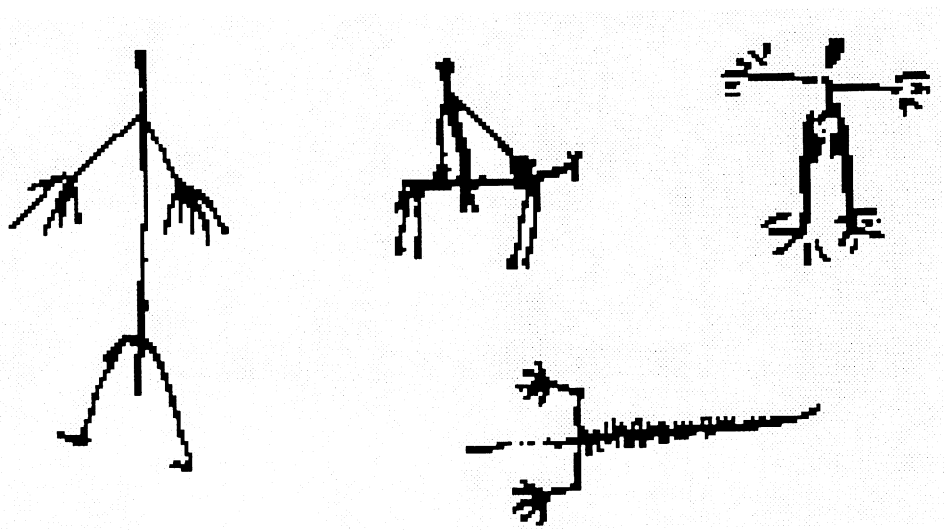

Figure 3: Drawings of a woman, a horse and a rider, a man and a crocodile made by Tallensi tribes (adapted from [Deregowski 1989]). See also [Marr and Nishihara 1978].

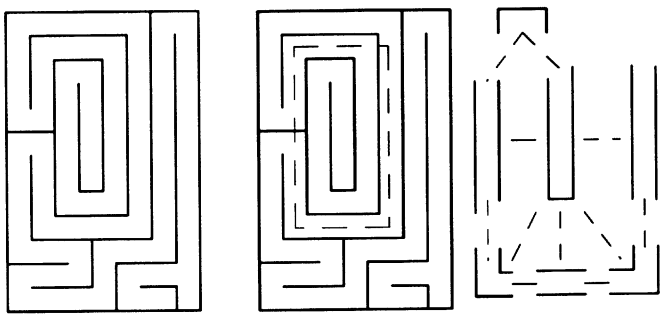

Figure 4: *Left:* Maze. *Center:* A portion of the maze is highlighted with a discontinuous doted line. *Right:* A candidate description of the highlighted region that may be useful to find an exit to the maze. In this paper, we are interested in finding frames or skeletons that can yield high-level descriptors like the one shown.

(the first problem above) in skeleton finding computations is the failure to tolerate noisy or circular shapes which often results in disconnected and distorted frames. A notched rectangle is generally used to illustrate this point, see [Serra 1982], [Brady and Connell 1987] or [Bagley 1985] for some more examples. [Heide 1984], [Bagley

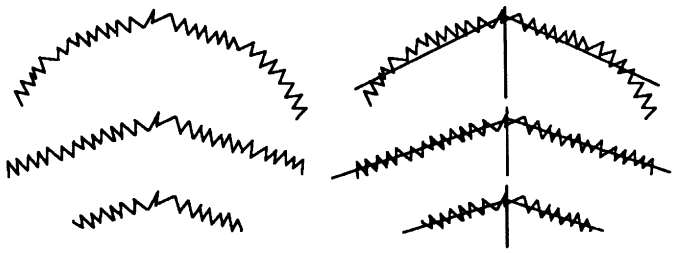

Figure 5: Finding corners is hard because they depend on scale. Here we present compelling evidence adapted from [Lowe 88]. The scheme presented in this paper can locate corners of this type, independently of scale, because it looks for the largest possible scale.

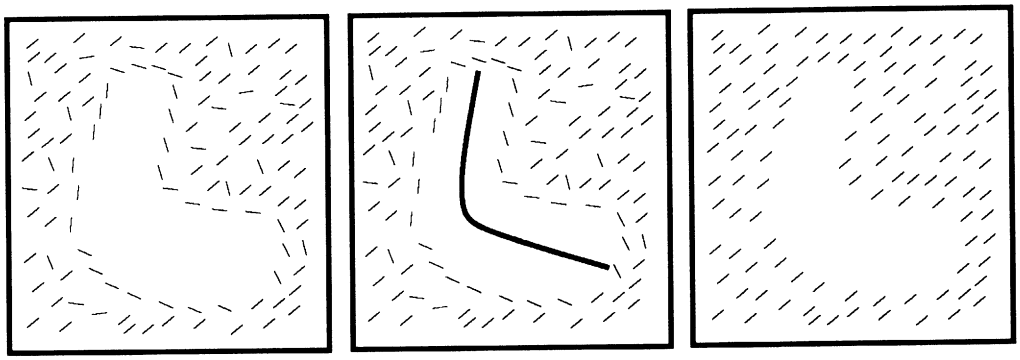

Figure 6: Finding the bent blob in the *left* image would be easy if we had the bent frame shown in the *center*. *Right:* Another blob defined by orientation elements of a single orientation. The scheme presented in this paper needs some modifications before it can attempt to segment the blob on the right (see text).

1985], [Brady and Connell 1987], [Fleck 1985, 1986, 1988], [Fleck 1989] suggest to solve this stability problem by working on the obtained SLS eliminating the portions of it that are due to noise, connecting segments that come from adjacent parts of the shape and by smoothing the contours at different scales. In our scheme, symmetry gaps are closed automatically we look for the largest scale available in the image and the frame depends on all the contour, not just a small portion making the scheme robust to small changes in the shape.

SAT and SLS are bad for circular shapes. [Fleck 86] addressed this problem by designing a separate computation to handle circular shapes, the Local Rotational Symmetries. Our scheme has a preference for the vertical that will bias the frame

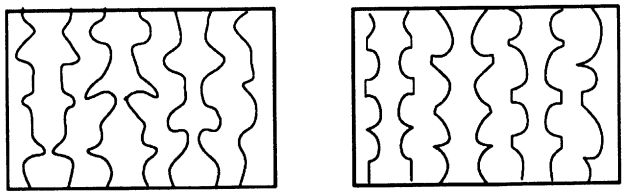

Figure 7: This Figure illustrates the importance of symmetry and convexity in grouping. The curves in the left image are grouped together based on symmetry. On the right image, convexity overrides symmetry, after [Kanizsa and Gerbino 76]. This grouping can be performed with the network presented in this paper by looking for the salient axes in the image.

Figure 8: All the shapes in this Figure have been drawn by adding a small segment to the shape in the middle. At a first glance, all of these shapes would be interpreted as two blobs. But if we are told that they are letters then finer distinctions are made between them. When we use such high level knowledge we perceive this shapes as being different and therefore their associated skeletons would differ dramatically.

towards a vertical line in circular shapes. When the shape is composed of a long straight body attached to a circular one (e.g. a spoon) then the bias will be towards having only one long axis in the direction of the long body.

3 Inertia Surfaces and Tolerated Length

If we are willing to restrict the frame to a single straight line then the axis of least inertia is a good reference frame because it provides a connected skeleton and it can handle nonsymmetric connected shapes. The inertia $\mathbf{In}(SL, A)$ of a shape A with respect to a straight line SL is defined as (See Figure 10):

$$\mathbf{In}(SL, A) = \int_A \mathcal{D}(a, SL)^2 da \tag{1}$$

The integral is extended over all the area of the shape, and $\mathcal{D}(a, SL)$ denotes the distance from a point a of the shape to the line SL. The axis of least inertia of a

shape A is defined as the straight line SL minimizing $\mathbf{In}(SL, A)$.

A naive way of extending the definition of axis of least inertia to handle bent curves would be to use Equation 1, so that the skeleton be defined as the curve C minimizing $\mathbf{In}(C, A)$. This definition is not useful if C can be any arbitrary curve because a highly bent curve that goes through all points inside the shape would have zero inertia (see Figure 11). There are two possible ways to avoid this problem: either we define a new measure that penalizes such curves or we restrict the set of permissible curves. We chose the former approach and we call the new measure defined in this paper (see equation 4) the *inertia*, the *skeleton saliency* or *saliency* of the curve. The skeleton saliency of a curve will depend on two local measures: the *inertia value* $\mathcal{I}$ that will play a role similar to that of $\mathcal{D}(p, a)$ in equation 1 and the *tolerated length* $\mathcal{T}$ that will prevent non-smooth curves from receiving optimal saliency values. The saliency of a curve will be defined for any curve C of length L starting at a given point p in the image. We define the problem as a maximization problem so that the "best" skeleton will be the curve that has the highest saliency value. By best we mean that the skeleton corresponds to the "most central curve" in the "most interesting (i.e. symmetric, convex, large)" portion of the image.

The inertia value

The inertia measure $\mathcal{I}$ for a point p and an orientation α is defined as (see Figure 12):

$$\mathcal{I}(p, \alpha) = 2R^{\frac{(R-r)^*}{R^*}},$$

Figure 12 shows how r, R and the inertia surfaces are defined for a given orientation α. $R = d(p_l, p_r)/2$ and $r = d(p, p_c)$, where p_l and p_r are the closest points of the contour that intersect with a straight line perpendicular to α (i.e. with orientation $\alpha + \pi/2$) that goes through p at opposite directions and p_c is the midpoint of the interval between these two points. For a given orientation, the inertia values of the points in the image form a surface that we call the *inertia surface* for that orientation. Figure 11 illustrates why the inertia values should depend on the orientation of the skeleton and Figure 13 shows the inertia surfaces for a square at four orientations.

Local maxima on the inertia values for one orientation indicate that the point is centered in the shape at that orientation. The absolute value of the local maximum indicates how large the section of the body is at that point for the given orientation, so that points in large sections of the body receive higher inertia values. The constant

s or *symmetry constant*, 2 in the actual implementation, controls the decrease in the inertia values for points away from the center of the corresponding section, the larger s is the larger the decrease. If s is very large only center points obtain high values and if $s = 0$ all points of a section receive the same value.

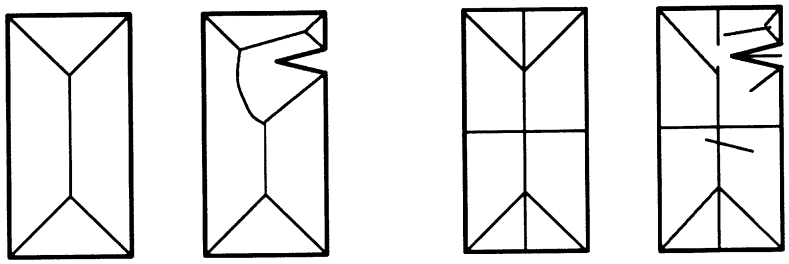

Figure 9: Skeletons found for a rectangle and a notched rectangle by SAT left and SLS right. Observe that the skeleton is highly distorted by the presence of the notch. The skeleton found by the computation presented in this paper on this shapes can be seen in Figures 7 and 8.

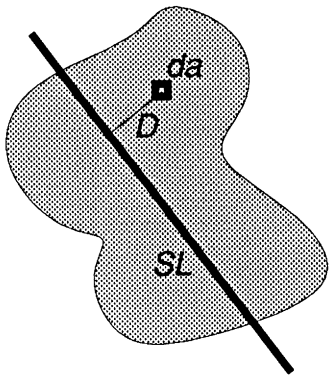

Figure 10: Definition of axis of inertia (see text for details).

The tolerated length

Figure 11 provides evidence that the curvature on a skeleton should depend on the width of the shape. As mentioned above, the *"tolerated length"* $\mathcal{T}$ will be used to evaluate the smoothness of a frame so that the curvature that is *tolerated* depends on the width of the section allowing high curvature only on thin sections of the shape.

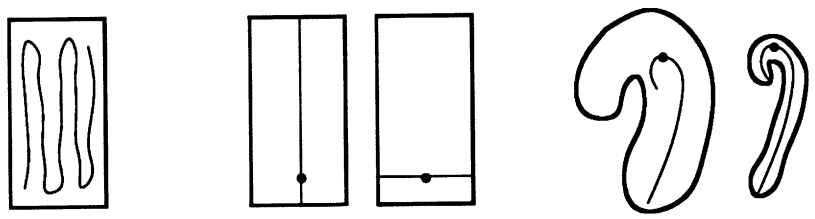

Figure 11: *Left:* A rectangle and a curve that would receive very low inertia according to Equation 1. *Center:* Evidence that the inertia value of a point should depend on orientation. *Right:* Evidence that the tolerated curvature on a skeleton should depend on the width of the shape.

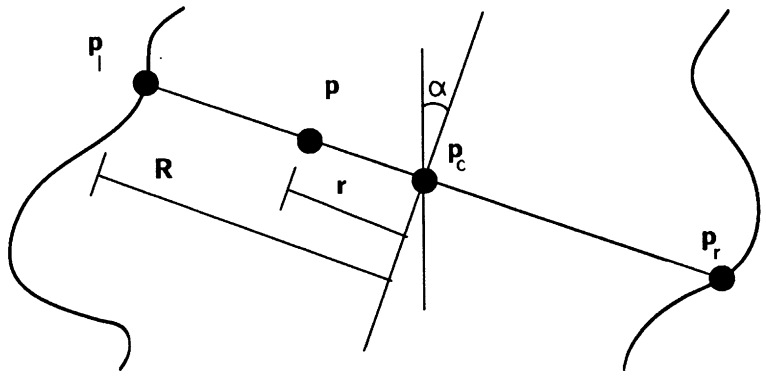

Figure 12: This Figure shows how the inertia surfaces are defined for a given orientation α. The value for the surface at a point p is $\mathcal{I}(R, r)$. The function $\mathcal{I}$ or *inertia function* is defined in the text. $R = d(p_l, p_r)/2$ and $r = d(p, p_c)$, where p_l and p_r are the points of the contour that intersect with a straight line perpendicular to α that goes through p at opposite directions and p_c is the midpoint of the interval between these two points. If there is more than one intersection along one direction then we use the nearest one. If there is no intersection at all then we give a preassigned value to the surface, 0 in the current implementation.

The saliency of a curve will be the sum of the inertia values "up to" the tolerated length so that for a high tolerated length, i.e. low curvature, the sum will include more terms and will be higher. The objective is that a curve that bends into itself within a section of the shape have a point within the curve with 0 tolerated length so that the saliency of the curve will not depend on the shape of the curve beyond that point. In other words, $\mathcal{T}$ should be 0 when the radius of curvature of the "potential" skeleton is smaller than the width of the shape at that point and a positive value

otherwise (with an increasing magnitude the smoother the curve is).

We define the *tolerated length* $\mathcal{T}$ for a curvature of radius r_c as:

$$\mathcal{T}(p,\alpha,r_c) = \begin{cases} 0 & \text{if } r_c < R + r \\ r_c(\pi - arccos(\frac{r_c-(R+r)}{r_c})) & \text{otherwise} \end{cases}$$

If a curve has a point with a radius of curvature r_c smaller than the width of the shape its tolerated length will be 0 and this, as we will see, results in a non-optimal curve[1]

In this section we have introduced the inertia surfaces and the tolerated length. We will define a salient frame of reference to be a high and long curve in the inertia surfaces that is as smooth as possible based on the tolerated length. Our approach is to associate a measure to *any* curve in the plane and to find the one that yields the highest possible value. The inertia value will be used to ensure that curves close to the center of large portions of the shape receive high values. The tolerated length will be used to ensure that curves bending beyond the width of the shape receive low values. In the next section we will investigate how such a curve might be computed in a general framework and in section 5 we will see how to include the inertia values and the tolerated length in the computation and what is the definition of the saliency measure that results.

4 A Network to Find Salient Curves

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 section we will apply these algorithms to finding long and smooth ridges in the inertia surfaces. [Mahoney 87] 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 88] devised a saliency measure and a dynamic programming algorithm that can find such salient curves in a binary image. 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 Appendix I.

[1]Because of this, if a simply connected closed curve has a radius of curvature lying fully inside the curve then it will not be optimal. Unfortunately I have not been able to prove that any simply connected closed curve has such a point nor that there is a curve with such a point.

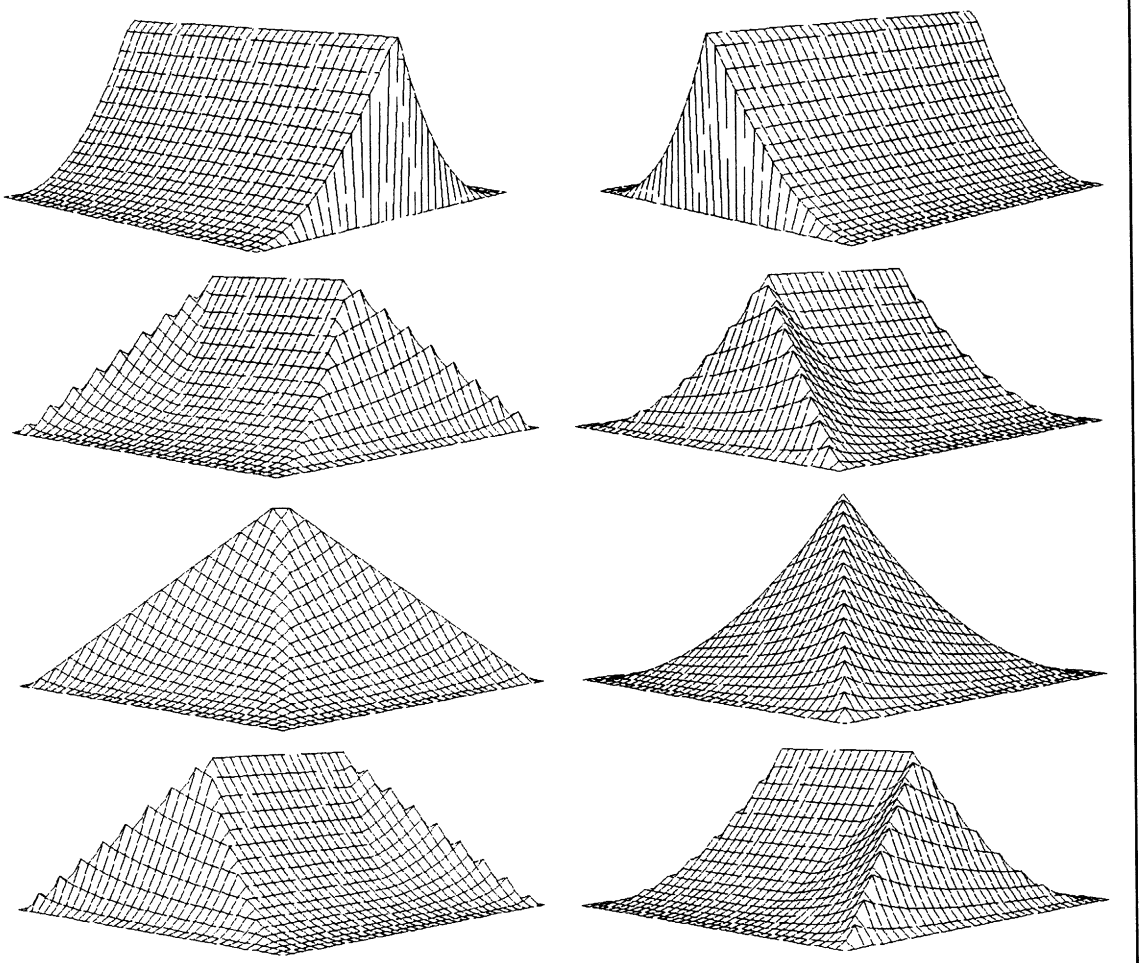

Figure 13: Plots of the inertia surfaces for a square for orientations parallel to the sides (left two plots) and parallel to the diagonals (right two plots).

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 Appendix I 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:

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

$$s_{i,j}(0) = \mathcal{F}(0, \boldsymbol{p}_e, \boldsymbol{p}_j, 0, 0) \tag{2}$$

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 i.e. iff

$$\forall \boldsymbol{p}, x, y \quad x < y \quad \longrightarrow \mathcal{F}(\boldsymbol{p}, x) < \mathcal{F}(\boldsymbol{p}, y), \tag{3}$$

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:

$$\mathcal{F}(\boldsymbol{p}, x) = f(\boldsymbol{p}) + g(x) * h(\boldsymbol{p}) \tag{4}$$

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 2).

5 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 Section 3. 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 4) are defined as:

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

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 5 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.

$$S_L = \int_0^L \mathcal{I}(l)\rho^{\int_0^l \frac{1}{\alpha\mathcal{T}(t)}dt}dl \tag{5}$$

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.

Smoothing

Straight lines that have an orientation different from one of the eight network orientations generate curvature impulses due to the discretization imposed on them, essentially 45 or 90 degrees (in a number of pixels, per unit length, which can be made arbitrarily large with a finer grid). This results in a reduction of the saliency for such curves biasing the network towards certain orientations. To prevent this, we made an implementation of the network that included a smoothing term that enables the processors to change their orientation at each iteration, instead of keeping only one of the eight initial orientations. At each iteration, the new orientation is computed by looking at those nearby pixels of the curve which lie on a straight line (so that curvature is minimized).

This allows greater flexibility but at the expense of breaking the optimization relation shown in Equation 3. A similar problem is encountered with the smoothing term suggested in [Sha'ashua and Ullman 1988].

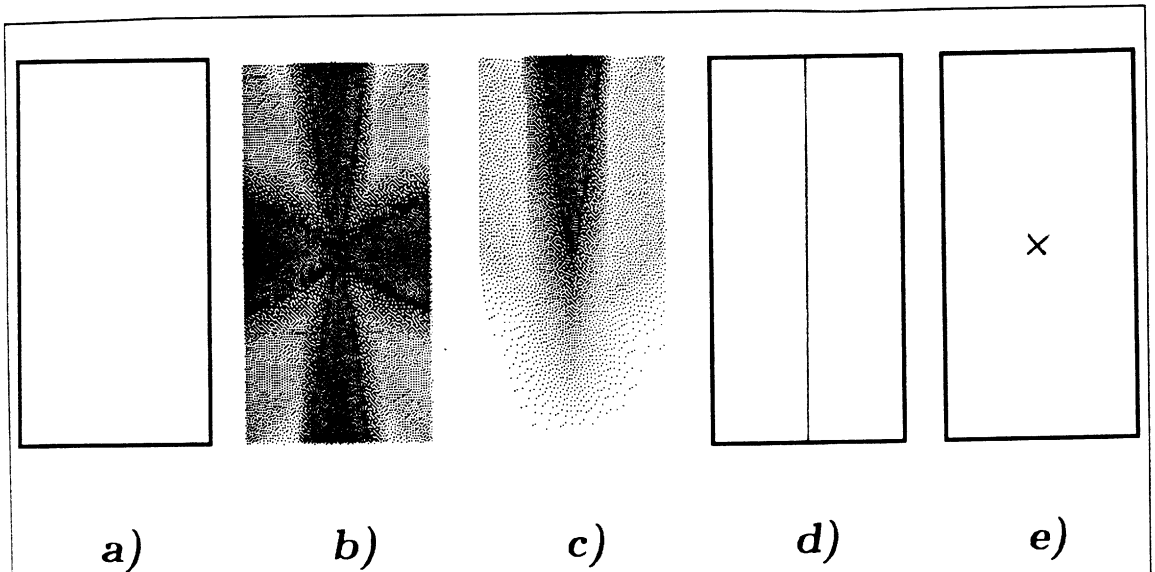

Figure 14: a) Rectangle. b) Skeleton sketch for the rectangle. Circles along the contour indicate local maxima in the skeleton sketch. c) Skeleton sketch for the rectangle for one particular orientation, vertical-down in this case. d) Most salient curve. e) Most interesting point for the most salient curve.

6 Results and Applications

In this section we will present some results and applications of the frame computation and in the next section we will discuss the connections of our findings to human perception.

The network described in the previous section has been implemented on a Connection Machine and tried on a variety of images. As mentioned above, the implementation works in two stages. First, the distance to the nearest point of the shape is computed at different orientations all over the image so that the *inertia surfaces* and the *tolerated length* can be computed, this requires a simple distance transform of the image. In the second stage, the network described in section 5 computes the saliency of the best curve starting at each point in the image for different orientations - eight in the current implementation. The number of iterations needed is bounded by the length of the most salient curve but in general a much smaller number of iterations will suffice. In all the examples shown in this paper the images were 128 by 128 pixels and 128 iterations were used. However, in most of the examples, the results do not change after about 40 iterations. In general, the number of iterations needed is bounded by the width of the shape measured in pixels.

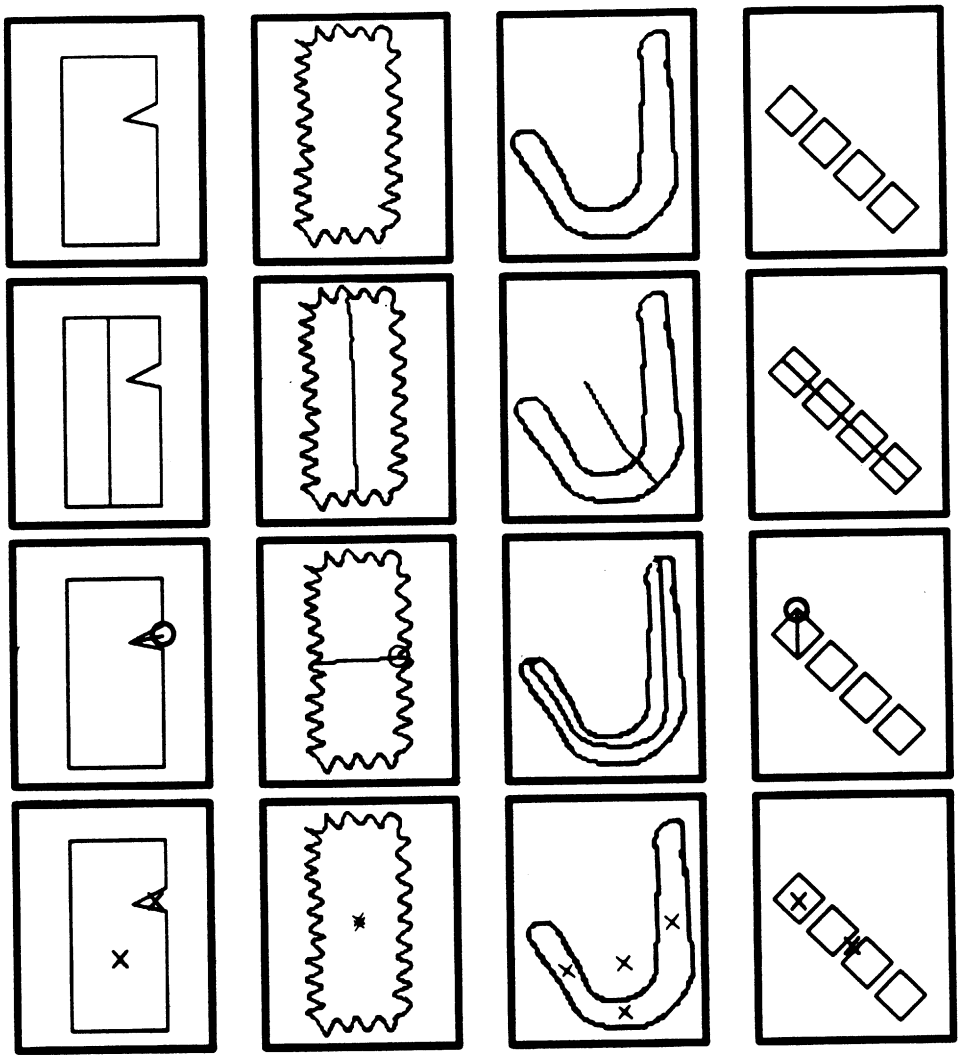

Figure 15: *Top:* Four shapes, a notched square, a stamp, a J, and Mach's demonstration. *Second row:* The most salient curve found by the network for each of them. Observe that the scheme is very stable under noisy or bent shapes. *Third row:* The most salient curve starting inside the shown circles. For the J shape the curve shown is the most salient curve that is inside the shape. *Fourth row:* The most interesting point according to the curves shown in the two previous rows. See text for details.

The skeleton sketch and the most salient curve

The *skeleton sketch* contains the saliency value for the most salient curve at each point. The skeleton sketch is similar to the saliency map described in [Sha'ashua and

Ullman 1988] and [Koch and Ullman 1985] because it provides a saliency measure at every point in the image. Figure 14 shows the skeleton sketch for a square. The best skeleton can be found by tracing the curve starting at the point having the highest skeleton saliency value. Figure 15 shows a few shapes and the most salient curve found by the network for each of them. Observe that the algorithm is very robust in the presence of non smooth contours. Given a region in the image we can find the best curve that starts in the region by finding the maxima of the skeleton sketch in the region, see Figure 15. In general, any local maximum in the skeleton sketch corresponds to a curve accounting for a symmetry in the image. Local maxima in the shape itself are particularly interesting since they correspond to features such as corners.

The most salient point

In many vision tasks, besides being interested in finding a salient skeleton, we are interested in finding a particular point related to the curve, shape or image. This can be due to a variety of reasons, because it defines a point in which to start subsequent processing to the curve or because it defines a particular place in which to shift our window of attention. Different points can be defined, the point with the highest saliency value is one of them, because it can locate relevant features such as corners.

Another interesting place in the image is the most central point in a curve which can be computed by our scheme by looking for the saliency values along the curve at both directions within the curve. The most central point can be defined as the point where these two values are "large and equal", the point that maximizes $\min(p_l, p_r)$ has been used in the current implementation, other functions are possible, see Figure 15 for some examples. Observe in Figure 15 that a given curve can have several *central points* due to different local maxima. This point can be used to direct future processing[2].

Similarly, the most central point in the image can be defined as the point that maximizes $\min(p_l, p_r)$ for all orientations.

[2]See also [Reisfeld, Wolfson and Yeshurun 1988] where a scheme to detect interest points was presented. Their scheme was scale dependent contrary to our scheme which selects the larger structure as the most interesting one, independently of the scale at which the scene is seen.

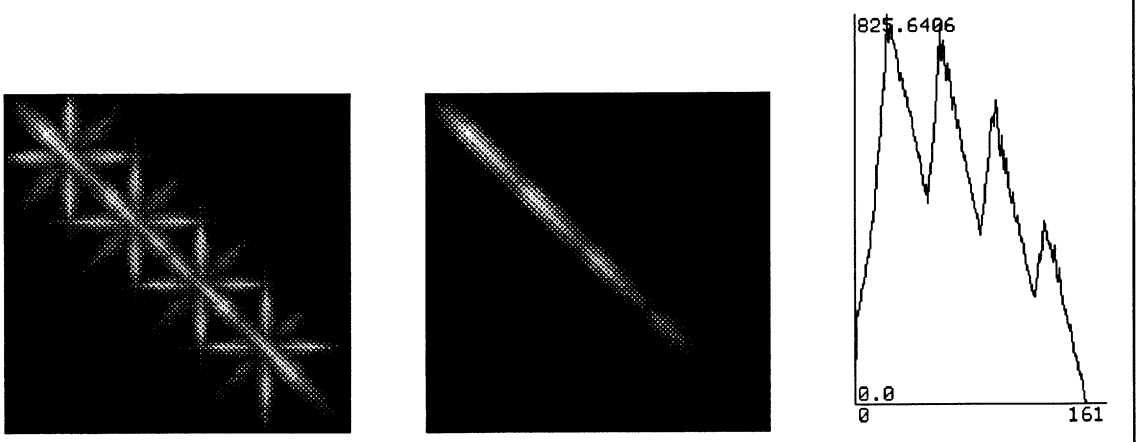

Figure 16: *Left:* Skeleton Sketch for Mach's demonstration (Original image in previous Figure top right). *Center:* Skeleton Sketch for *one orientation only. Right:* Slice of the "one-orientation" Skeleton Sketch through one of the diagonals of the image. Note that the values decrease across the gaps and increase inside the square (see also [Palmer and Bucher 1981]).

Shape description

Each locally salient curve in the image corresponds to a *symmetric region* in one portion of the scene. The selection of the set of most interesting frames corresponding to the different parts of the shape yields a part description of the scene. Doing this is not trivial (See [Shashua and Ullman 1990]) because a salient curve is surrounded by other curves of similar saliency. In general, a curve displaced one pixel to the side from the most salient curve will have a saliency value similar to that of the most salient one and higher than that of other locally most salient curves. In order to inhibit these curves we color out from a locally maximal curve in perpendicular directions to suppress parallel nearby curves. The amount to color can be determined by the average width of the curve. Once nearby curves have been suppressed we look for the next most salient curve and iterate this process. Another approach to find a group of several curves, not just one, is given in [Sha'ashua and Ullman 1990]. Both approaches suffer from the same problem: the groups obtained do not optimize a simple global maximization function.

Figure 17 shows the skeleton found for an airplane. The skeleton can then be used to find a part description of the shape in which each component of the frame has different elements associated describing it: a set of contours from the shape, a saliency measure reflecting the relevance or saliency that the component has within

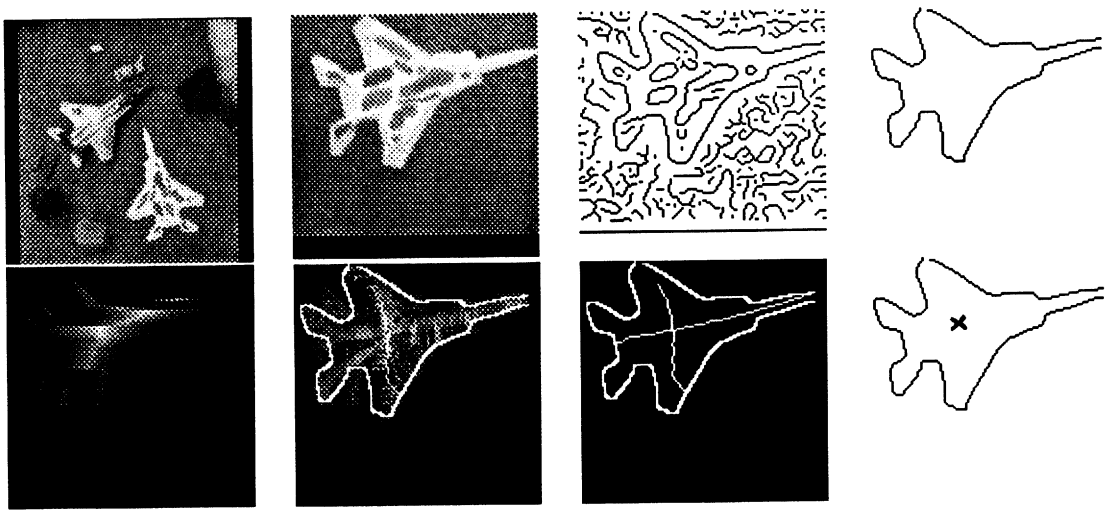

Figure 17: *Top:* Image; airplane portion enlarged; its edges; airplane without short edges. *Bottom:* Vertical inertia surface; skeleton sketch; skeleton; most salient point.

the shape, a central point, a location within the shape.

Inside-outside

The network can also be used to determine a continuous measure of inside-outside (see also [Subirana-Vilanova and Richards 1991]). The distance from a point to the frame can be used as a measure of how near the point is to the outside of the shape. This measure can be computed using a scheme similar to the one used to inhibit nearby curves as described in the previous paragraph: coloring out from the frame at perpendicular orientations, and using the time where a point is colored as a measure of how far from the frame the point is. The saliency of a curve provides a measure of the area swept by the curve which can be used to scale the coloring process.

7 Relation to human perception

The skeleton found by the network for a given shape corresponds roughly with the central regions of the shape. In this section we show how the scheme can handle various peculiarities of human perception.

Important frames of reference in the perception of shape and spatial relations by humans include: that of the perceived object, that of the perceiver and that of the environment. In this paper we have concentrated on the first. A considerable amount of effort has been devoted to study the effects of orientation of such a frame (relevant results include, to name but a few [Attneave 1967], [Shepard and Metzler 1971], [Rock 1973], [Cooper 1976], [Wiser 1980, 1981], [Schwartz 1981], [Shepard and Cooper 1982], [Jolicoeur and Landau 1984], [Jolicoeur 1985], [Palmer 1985], [Palmer and Hurwitz 1985], [Corballis and Cullen 86], [Maki 1986], [Jolicoeur, Snow and Murray 1987], [Parsons and Shimojo 1987], [Robertson, Palmer and Gomez 1987], [Rock and DiVita 1987], [Bethel-Fox and Shepard 1988] [Shepard and Metzler 1988], [Corballis 1988], [Palmer, Simone and Kube 1988], [Georgopoulos, Lurito, Petrides, Schwartz and Massey 1989], [Tarr and Pinker 1989]). Our scheme suggests a computational model of how such an orientation may be computed, i.e. the selected orientation is that of the most salient skeleton when it is restricted to be straight (α and ρ close to 0).

The influence of the environment on the frame has been extensively studied too [Mach 1914], [Attneave 1968], [Palmer 1980], [Palmer and Bucher 1981], [Humphreys 1983], [Palmer 1989]. In some cases the perception of the shape can be biased by the frame of the environment. In particular, humans have a bias for the vertical in shape description (see [Rock 73]) so that some shapes are perceived very differently depending on the orientation at which they are viewed, for example a rotated square is perceived as a diamond (see Figure 23). This bias can be taken into account in our scheme by adding some constant value to the inertia surface that corresponds to the vertical orientation so that vertical curves receive a higher saliency value. Adding the bias towards the vertical is also useful because it can handle non-elongated objects that are not symmetric, so that the preferred frame is a vertical axis going through the center of the shape[3].

In other cases, the preferred frame is defined by the combination of several otherwise non salient frames. This is the case in Mach's demonstration, first described by E. Mach at the beginning of this century (see Figure 15). Our scheme incorporates this behavior because the best curve can be extended beyond one object increasing the saliency of one axis by the presence of objects nearby, especially when the objects have salient aligned axis. This example also illustrates the tolerance of the scheme to fragmented shapes.

The shape of the frame has received very little attention. [Subirana-Vilanova 1990] proposed that in some cases, a curved frame might be useful (see also Figure 2 and [Palmer 1989]). In particular, he proposed to recognize elongated curved objects by

[3]As discussed in section 3, another alternative is to define a specific computation to handle the portions of the shapes that are circular [Fleck 86], [Brady and Scott 88].

unbending them using their main curved axis as a frame to match the unbended versions. [Subirana-Vilanova and Richards 1991] have shown that such strategy is not always used by the human visual system.

In figure-ground segregation, reference frame computation and perceptual organization it is well known that humans prefer symmetric regions over those that are not (see Figures 7 and references above[4]). Symmetric regions can be discerned in our scheme by looking for the points in the image with higher skeleton saliency values. However, [Kanisza and Gerbino 76] have shown that in some cases convexity may override symmetry (see Figure 7). Convexity information can be introduced in the inertia surfaces by looking at the distances to the shape and at the convexity at these points so that frames inside a convex region receive a higher symmetry value. Observe that the relevant scale of the convexity at each point can be determined by the distances to the shape R and r.

The location of the frame of reference [Richards and Kaufman 1969], [Kaufman and Richards 1969], [Carpenter and Just 1978], [Cavanagh 1978, 1985], [Palmer 1983], [Nazir and O'Reagan 1990] is related to attention and eye movements [Yarbus 1967] and influences figure-ground relations (e.g. Figure D9 in [Shepard 1990]). We have shown how certain salient structures and individual points can be selected in the image using the Skeleton Sketch. Subsequent processing stages can be applied selectively to the selected structures, endowing the system with a capacity similar to the use of selective attention in human vision. The points provided by the Saliency Sketch are in locations central to some structures of the image and could guide processing in a way similar to the direction of gaze in humans (e.g. [Yarbus 1967]).

[Palmer 1983] studied the influence of symmetry on figural goodness. He computed a "mean goodness rating" associated to each point inside a figure. For a square (see Figure 4 in [Palmer 1983]), he found a distribution similar to that of the skeleton sketch shown in Figure 14. The role of this measure is unclear but our scheme suggests that it can be computed bottom-up and hence play a role in the recognition of the shape.

Perhaps, this measure is involved in providing translation invariance so that objects are first transformed into a canonical position. This suggestion is similar to others that attempt to explain rotation invariance (see references above) and it could be tested in a similar way. For example, one can compute the time to learn/recognize an object (from a class sharing a similar property such as the one shown in Figure 20) in terms of a given displacement in fixation point (or orientation in the references

[4]The role of symmetry has been studied also for random dot displays [Barlow and Reeves 1979], [Barlow 1982] and occlusion [Rock 84].

above).

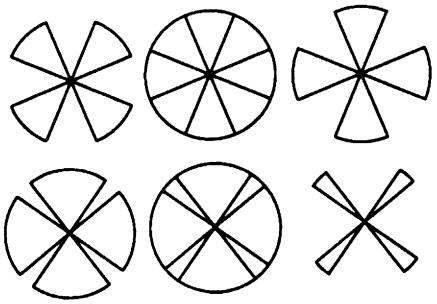

Figure 18: *Top Center:* Figure is often seen as shown on the *right*, (and ground as *on the left*) due to vertical bias. *Bottom Center:* Preference for the vertical, and preference for large objects is over-ridden here by the preference for small structures (after [Rock 1985]). The network presented in this paper would find the *left* object as figure due to its preference for large structures. Further research is necessary to clarify when small structures are more salient.

Figure 19: Like in the previous Figure, small structures define the object depicted in this image. Drawing from Miró. This image would confuse the network presented in [Sha'ashua and Ullman 1988].

8 What's New

In this paper we have presented C.I.F. (Curved Inertia Frames), a novel scheme to compute curved symmetry axes. Previous schemes either use global information, but compute only straight axes, or compute curved axes and use only local information.

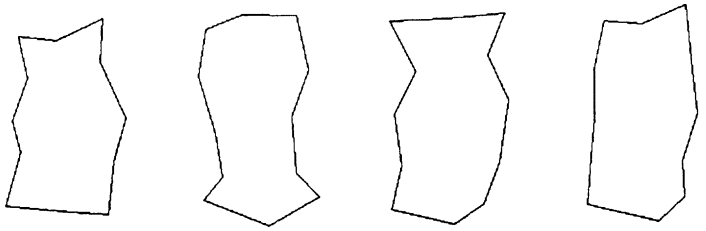

Figure 20: Does the time to recognize/learn this objects depend on the fixation point? (See text for details.)

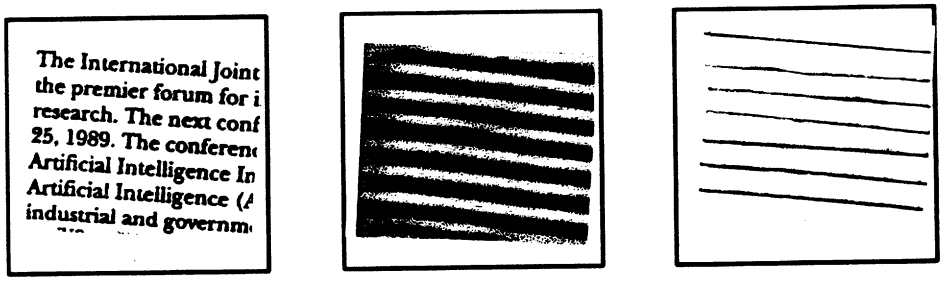

Figure 21: *Left:* Text image. *Center:* Output of the convolution of the text image with an elongated horizontal gabor filter. *Right:* Most salient curves.

The scheme presented in this paper can extract curved symmetry axes and use global information. This gives the scheme some clear advantages over previous ones, such as: 1) it can compute curved axes, 2) it provides connected axes, 3) it is remarkably stable to changes in the shape, 4) it provides a measure associated with the relevance of the axes in the shape, which can be used for shape description and for grouping based on symmetry and convexity 5) it can tolerate noisy and spurious data 6) it provides central points of the shape.

We have suggested a novel scheme to recognize elongated flexible objects by "unbending" them using C.I.F. and demonstrated the "unbending" transformation on the simple shapes shown in Figure 2. This is useful because flexible objects can be matched as rigid ones once they have been transformed to the canonical straight orientation. In fact, the canonical orientation need not be straight. If the objects generally deviate from a circular arc then the canonical representation could store the object with a circular principal axis.

We believe that the Skeleton Sketch and its associated curves and interest points

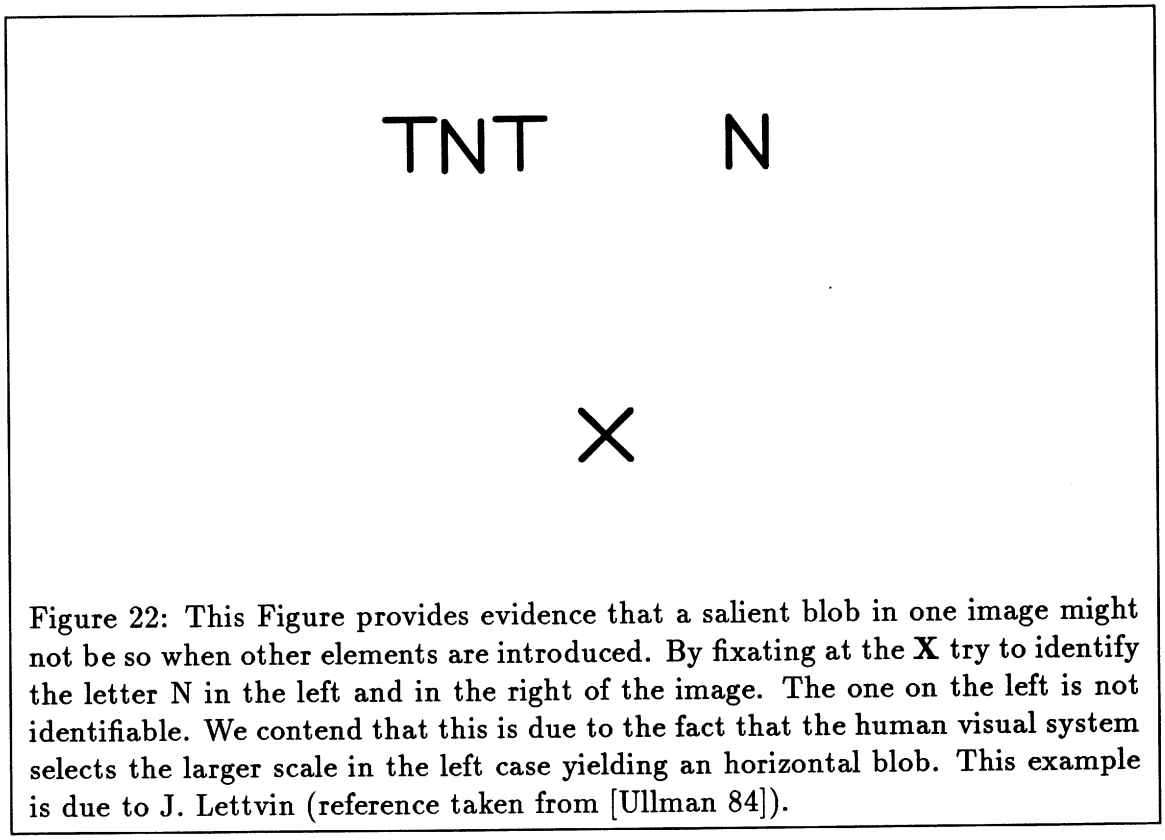

Figure 22: This Figure provides evidence that a salient blob in one image might not be so when other elements are introduced. By fixating at the **X** try to identify the letter N in the left and in the right of the image. The one on the left is not identifiable. We contend that this is due to the fact that the human visual system selects the larger scale in the left case yielding an horizontal blob. This example is due to J. Lettvin (reference taken from [Ullman 84]).

can also be used for part segmentation, attention, figure-ground segmentation, perceptual organization, recognition and feature detection. However, further research is necessary to support this.

The Skeleton Sketch suggests a way in which interest points can be computed bottom-up, and hence that they might be useful as anchor structures for aligning model to object. It also provides a continuous measure that can be used to determine the distance from the center of the object, suggesting a number of experiments. For example, one could test whether the time to learn/recognize an object depends on the fixation point in a similar way in which a dependence has been found in human perception between object orientation and recognition time/accuracy (see references above). This could be done on a set of similar objects of the type shown in Figure 20.

We have introduced the inertia surfaces and the tolerated length and we have shown how they can be used to find skeletons using a sophisticated version of an algorithm presented previously [Sha'ashua and Ullman 88]. In the Appendix we show some limitations on the functions that can be optimized using such algorithm. Similar measures might be used to find skeletons using other algorithms such as those

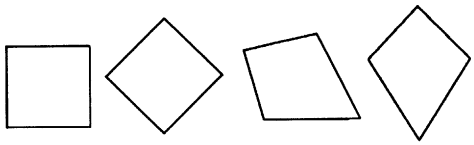

Figure 23: A square has four symmetry axis all of which could potentially be used to describe it. Depending which one of them is chosen this shape appears as a square or as a diamond. This suggests that when there is ambiguity the vertical can play an important role. The two trapezoids, on the right further illustrate that even when a shape has several symmetry axis the vertical might be preferred even if it does not correspond to a perfect symmetry axis. Observe that the vertical might be overridden by an exterior frame which can be defined by the combination of several otherwise not salient frames from different shapes such as Mach demonstration, see Figure 8.

presented in [Kass, Witkin and Terzopoulos 88] and [Zucker, Dobbins and Iverson 89].

The network presented in this paper computes skeletons in 2 dimensional images. The network can be extended to finding 3 dimensional skeletons from 3 dimensional data since the local estimates for orientation and curvature can be found in a similar way and the network extends to 3 dimensions - this, of course, at the cost of increasing the number of processors. The problem of finding 3D skeletons from 2D images is more complex; however, in most cases the projection of the 3D skeleton can be found by working on the 2D projection of the shape, especially for elongated objects (see shapes in [Snodgrass and Vanderwart 1980]).

The scheme presented in this paper has two important limitations. First, it relies on discontinuities. This can be overcome, to a certain extent, by extending the scheme to finding high, long and smooth curves in arbitrary surfaces (but see [Subirana-Vilanova and Sung 1992]). The scheme, as presented here, searches for the *best curve* using local estimates for orientation and curvature. The estimates can be obtained in an arbitrary surface by convolving it with oriented gabor filters at different orientations and scales. This could be applied to many tasks in vision. An example of such applications is finding dark blobs in images (see figure 21), the scheme selects both a region and a scale in the image.

The second limitation is that it has a bias for large structures. This is generally a good rule, even for human perception, except in some cases, see Figures 19, 18 and 22. The example of Figure 18 provides evidence that the preference for small

objects can not be only due to pop-out effects. A naive solution would be to rank the scale of the different contours in the image and find the salient ones in terms of their location in such ordering. This distinction had not been made clear before and deserves further treatment.

Acknowledgments

Thanks to Ronen Basri, David Beymer, Thomas Breuel, Shimon Edelman, Eric Grimson, Joe Heel, Tomaso Poggio, Amnon Sha'ashua, Shimon Ullman and Woody Yang for useful discussions and suggestions for improving the presentation of the paper. Thanks to James Mahoney, Eric Saund and the staff of the Electronics Documents Laboratory at Xerox P.A.R.C. where I did the Connection Machine implementation of the scheme presented in this paper during summer 1989.

Appendix I

In the appendix we show that the set of possible saliency measures that can be computed with the network defined in [Sha'ashua and Ullman 88] (see also section 4) is limited.

Proposition 1 *The use of more than one state variable in the saliency network defined in section 4 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 section 4. We will do the proof for the case of two state variables, the generalization of the proof to more state variables follows naturally. Each edge will have 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. □.

References

[1] C. Arcelli, L.P. Cordella, and S. Levialdi. From local maxima to connected skeletons. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-3(2):134–143, 1981.

[2] H. Asada and M. Brady. The curvature primal sketch. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(1):2–14, 1985.

[3] F. Attneave. Triangles as ambiguous figures. *American Journal of Psychology*, 81:447–453, 1968.

[4] F. Attneave and R.K. Olson. Discriminability of stimuli varying in physical and retinal orientation. *Journal of Experimental Psychology: Human Perception and Performance*, 74:149–157, 1967.

[5] S.C. Bagley. Using models and axes of symmetry to describe two-dimensional polygonal shapes. Master's thesis, Massachusetts Institute of Technology, 1985.

[6] H.B. Barlow. The past, present and future of feature detectors. In D.G. Albrecht, editor, *Recognition of pattern and form. Proceedings of a conference held at the University of Texas at Austin, 1979*, pages 4–32. Springer-Verlag, 1982.

[7] H.B. Barlow and B.C. Reeves. The versatility and absolute efficiency of detecting mirror symmetry in random dot displays. *Vision Research*, 19:783–793, 1979.

[8] C.E. Bethell-Fox and R.N. Shepard. Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance*, 1:12–23, 1988.

[9] I. Biederman. Human image understanding: Recent research and a theory. *Computer Vision, Graphics, and Image Processing*, 32:29–73, 1985.

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

[11] H. Blum. A transformation for extracting new descriptors of shape. In Walthen Dunn, editor, *Models for the perception of speech and visual form*, pages 362–380. The MIT Press, Cambridge, MA, 1967.

[12] H. Blum and R.N. Nagel. Shape description using weighted symmetric axis features. *Pattern Recognition*, 10:167–180, 1978.

[13] R.C. Bolles and R.A. Cain. Recognizing and locating partially visible objects: The local-feature-focus method. *International Journal of Robotics Research*, 1(3):57–82, 1982.

[14] G. Borgefors. Distance transformations in digital images. *Computer Vision, Graphics, and Image Processing*, 34:344–371, 1986.

[15] M. Brady. Criteria for the representations of shape. In J. Beck, B. Hope, and A. Rosenfeld, editors, *Human and Machine Vision*, pages 39–84. Academic Press, New York, 1983.

[16] M. Brady and H. Asada. Smoothed local symmetries and their implementation. *International Journal of Robotics Research*, 3(3):36–61, 1984.

[17] M. Brady and G. Scott. Parallel algorithms for shape representation. In Ian Page, editor, *Parallel Architectures and Computer Vision*. OUP, 1988.

[18] M. Brady and A. Yuille. An extremum principle for shape from contour. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(3):288–301, 1984.

[19] C. Broit. *Optimal registration of deformed images*. PhD thesis, University of Pennsylvania, 1981.

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

[21] P.A. Carpenter and M.A. Just. Eye fixations during mental rotation. In J.W. Senders, D.F. Fisher, and R.A.Monty, editors, *Eye movements and the higher psychological functions*, pages 115–133. Erlbaum, Hillsdale, NJ, 1978.

[22] P. Cavanagh. Size and position invariance in the visual system. *Perception*, 7:167–177, 1978.

[23] P. Cavanagh. Local log polar frequency analysis in the striate cortex as a basis for size and orientation invariance. In D. Rose and V. Dobson, editors, *Models of the visual cortex*, pages 85–95. Wiley, 1985.

[24] J.H. Connell. Learning shape descriptions: generating and generalizing models of visual objects. Technical Report AI TR No. 853, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1985.

[25] J.H. Connell and M. Brady. Generating and generalizing models of visual objects. *Artificial Intelligence*, 31:159–183, 1987.

[26] L.A. Cooper. Demonstration of a mental analog to an external rotation. *Percepion and Psychophysics*, 1:20–43, 1976.

[27] M.C. Corbalis. Recognition of disoriented shapes. *Psychological Review*, 95:115–123, 1988.

[28] M.C. Corbalis and S Cullen. Decisions about the axes of disoriented shapes. *Mem. Cognition*, 14:27–38, 1986.

[29] J.B. Deregowski. Real space and represented space: Cross-cultural perspectives. *Behavioral and Brain Sciences*, 12:51–119, 1989.

[30] M.M. Fleck. Local rotational symmetries. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, pages 332–337, 1986.

[31] M.M. Fleck. Boundaries and topological algorithms. Technical Report AI-TR-1065, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1988.

[32] A.P. Georgopoulos, J.T. Lurito, M. Petrides, A.B. Schwartz, and J.T. Massey. Mental rotation of the neuronal population vector. *Science*, 243:234–236, 1989.

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

[34] S.S. Heide. A hierarchical representation of shape from smoothed local symmetries. Master's thesis, Massachusetts Institute of Technology, 1984.

[35] D.D. Hoffman and W.A. Richards. Parts of recognition. In S. Pinker, editor, *Visual Cognition*, pages 2–96. The MIT Press, Cambridge, MA, 1984.

[36] J. Hollerbach. Hierarchical shape description of objects by selection and modification of prototypes. Technical Report AI TR No. 346, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1975.

[37] G.W. Humphreys. Reference frames and shape perception. *Cognitive Psychology*, 15:151–196, 1983.

[38] D.W. Jacobs. The use of grouping in visual object recognition. A.I. Technical Report No. 1023, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, 1988.

[39] P. Jolicoeur. The time to name disoriented natural objects. *Mem. Cognition*, 13(4):289–303, 1985.

[40] P. Jolicoeur. A size-congruency effect in memory for visual shape. *Mem. Cognition*, 15(6):531–543, 1987.

[41] P. Jolicoeur and D. Besner. Additivity and interaction between size ratio and response category in the comparison of size-discrepant shapes. *Journal of Experimental Psychology: Human Perception and Performance*, 13(3):478–487, 1987.

[42] P. Jolicoeur and S.M. Kosslyn. Is time to scan visual images due to demand characteristics. *Mem. Cognition*, 13(4):320–332, 1985.

[43] P. Jolicoeur and M.J. Landau. Effects of orientation on the identification of simple visual patterns. *Canadian Journal of Psychology*, 38(1):80–93, 1984.

[44] P. Jolicoeur, D. Snow, and J. Murray. The time to identify disoriented letters: Effects of practice and font. *Canadian Journal of Psychology*, 41(3):303–316, 1987.

[45] G. Kanizsa and W. Gerbino. Convexity and symmetry in figure-ground organization. In M. Hele, editor, *Vision and Artifact*. Springer, New York, 1976.

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

[47] L. Kaufman and W. Richards. Spontaneous fixation tendencies for visual forms. *Percepion and Psychophysics*, 5(2):85–88, 1969.

[48] C. Koch and S. Ullman. Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, 4:219–227, 1985.

[49] A. Larsen and C. Bundesen. Size scaling in visual pattern recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 4(1):1–20, 1978.

[50] D.G. Lowe. *Perceptual Organization and Visual Recognition*. Kluwer Academic Publishers, Boston, 1986.

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

[52] E. Mach. *The analysis of sensations.* Chicago: Open Court, 1914.

[53] 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.

[54] R. Maki. Naming and locating the tops of rotated pictures. *Canadian Journal of Psychology*, 40(1):368–387, 1986.

[55] G. Marola. On the detection of the axis of symmetry of symmetric and almost symmetric planar images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-11(1):104–108, 1989.

[56] D. Marr. *Vision.* Freeman, San Francisco, CA, 1982.

[57] D. Marr and H.K. Nishihara. Representation and recognition of the spatial organization of three dimensional structure. *Proceedings of the Royal Society of London B*, 200:269–294, 1978.

[58] T.A. Nazir and J.K. O'Reagan. Some results on translation invariance in the human visual system. *Spatial Vision*, 5(2):81–100, 1990.

[59] R. Nevatia and T.O. Binford. Description and recognition of curved objects. *Artificial Intelligence*, 8:77–98, 1977.

[60] S.E. Palmer. Structural aspects of visual similarity. *Mem. Cognition*, 6(2):91–97, 1978.

[61] S.E. Palmer. What makes triangles point: Local and global effects in configurations of ambiguous triangles. *Cognitive Psychology*, 12:285–305, 1980.

[62] S.E. Palmer. On goodness, gestalt, groups, and garner. *Paper presented at Annual Meeting of Psychonomic Society. San Diego, California. Unpublished manuscript*, 1983.

[63] S.E. Palmer. The role of symmetry in shape perception. *Acta Psychologica*, 59:67–90, 1985.

[64] S.E. Palmer. Reference frames in the perception of shape and orientation. In B.E. Shepp and S. Ballesteros, editors, *Object perception: Structure and process*, pages 121–163. Hillsdale, NJ: Lawrence Erlbaum Associates, 1989.

[65] S.E. Palmer and N.M. Bucher. Configural effects in perceived pointing of ambiguous triangles. *Journal of Experimental Psychology: Human Perception and Performance*, 7:88–114, 1981.

[66] S.E. Palmer, E. Simone, and P. Kube. Reference frame effects on shape perception in two versus three dimensions. *Perception*, 17:147–163, 1988.

[67] L.M. Parsons and S. Shimojo. Percieved spatial organitzation of cutaneous patterns on surfaces of the human body in various positions. *Journal of Experimental Psychology: Human Perception and Performance*, 13(3):488–504, 1987.

[68] T. Pavlidis. A vectorizer and feature extractor for document recognition. *Computer Vision, Graphics, and Image Processing*, 35:111–127, 1986.

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

[70] D. Reisfeld, H. Wolfson, and Y. Yeshurun. Detection of interest points using symmetry. In IEEE Computer Society, editor, *Third International Conference on Computer Vision*, pages 62–65, 1990.

[71] W. Richards and L. Kaufman. "center-of-gravity" tendencies for fixations and flow patterns. *Percepion and Psychophysics*, 5(2):81–84, 1969.

[72] L.C. Roberson, S.E. Palmer, and L.M. Gomez. Reference frames in mental rotation. *Journal of Experimental Psychology: Human Perception and Performance*, 13(3):368–379, 1987.

[73] I. Rock. *Orientation and Form.* Academic Press, New York, 1973.

[74] I. Rock. *The logic of perception.* The MIT Press, Cambridge, MA, 1983.

[75] I. Rock and J. DiVita. A case of viewer-centered object perception. *Cognitive Psychology*, 19:280–293, 1987.

[76] A. Rosenfeld and J.L. Pfaltz. Distance functions on digital pictures. *Pattern Recognition*, 1:33–61, 1968.

[77] E. Saund. The role of knowledge in visual shape representation. A.I. Technical Report No. 1092, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1988.

[78] G.L. Scott, S.C. Turner, and A. Zisserman. Using a mixed wave-diffusion process to elicit the symmetry set. *Image and Vision Computing*, 7(1):63–70, 1989. This paper appeared also in the Procedings of the 4th Alvey Vision Conference, Manchester, England, 1988.

[79] R. Sekuler and D. Nash. Speed of size scaling in human vision. *Psychonomic Science*, 27:93–94, 1972.

[80] J. Serra. *Image Analysis And Mathematical Morphology.* Academic Press Inc., London, 1982.

[81] 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.

[82] A. Sha'ashua and S. Ullman. Grouping contours by iterated pairing network. In *NIPS*, 1990.

[83] R.N. Shepard. *Mind sights: Original visual illusions, ambiguities, and other anomalies, with a commentary on the play of mind in perception and art.* W.H. Freeman and Company, New York, 1990.

[84] R.N. Shepard and L.A. Cooper. *Mental Images and their Transformations.* The MIT Press, Cambridge, MA, 1982.

[85] R.N. Shepard and S. Hurwitz. Upward direction, mental rotation, and discrimination of left and right turns in maps. *Cognition*, 18:161–193, 1985.

[86] S. Shepard and D. Metzler. Mental rotation of three dimensional objects. *Science*, 171:701–703, 1971.

[87] S. Shepard and D. Metzler. Mental rotation: Effects of dimensionality of objects and type of task. *Journal of Experimental Psychology: Human Perception and Performance*, 14(1):3–11, 1988.

[88] S.P. Shwartz. The perception of disoriented complex objects. In *Proceedings of the 3rd Conference on Cognitive Sciences*, pages 181–183, Berkeley, 1981.

[89] 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.

[90] 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.

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

[92] J.B. Subirana-Vilanova and W. Richards. Perceptual organization, figure-ground, attention and saliency. A.I. Memo No. 1218, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1991.

[93] J.B. Subirana-Vilanova and K.K. Sung. Perceptual organization without edges. In *Proceedings Image Understanding Workshop*. Morgan and Kaufman, 1992.

[94] M. Tarr and S. Pinker. Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, 21:233–282, 1989.

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

[96] L.M. Vaina and S.D. Zlateva. The largest convex patches: A boundary-based method for obtaining parts. *Biological Cybernetics*, 62:225–236, 1990.

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

[98] M.A. Wiser. *The role of intrinsic axes in the mental representation of shapes.* PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 1980. See also the *Proceedings of 3rd Conference on Cognitive Sciences, Berkeley*, pp. 184-186, 1981.

[99] A.L. Yarbus. *Eye movements and Vision.* Plenum, New York, 1967.

[100] 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.

This blank page was inserted to preserve pagination.

CS-TR Scanning Project
Document Control Form

Date : 1/19/95

Report # AIM-1137

Each of the following should be identified by a checkmark:
Originating Department:

☒ Artificial Intellegence Laboratory (AI)
☐ Laboratory for Computer Science (LCS)

Document Type:

☐ Technical Report (TR) ☒ Technical Memo (TM)
☐ Other:_____

Document Information

Number of pages: 37(41-IMAGES)
Not to include DOD forms, printer intstructions, etc... original pages only.

Originals are:
☒ Single-sided or
☐ Double-sided

Intended to be printed as :
☐ Single-sided or
☒ Double-sided

Print type:
☐ Typewriter ☐ Offset Press ☒ Laser Print
☐ InkJet Printer ☐ Unknown ☐ Other:_____

Check each if included with document:

☐ DOD Form ☐ Funding Agent Form ☐ Cover Page
☐ Spine ☐ Printers Notes ☐ Photo negatives
☐ Other: _____

Page Data:

Blank Pages(by page number):_____

Photographs/Tonal Material (by page number): 5, 9, 17, 20, 21

Other (note description/page number):

Description :	Page Number:
IMAGE MAP - (1) UNNUMBERED TITLE PAGE	
(2-37) PAGES #'ED 1-36	
(38) SCANCONTROL	
(39-41) TRGT'S	

Scanning Agent Signoff:
Date Received: 1/19/95 Date Scanned: 1/24/95 Date Returned: 1/26/95

Scanning Agent Signature: Michael W. Cook

Scanning Agent Identification Target

Scanning of this document was supported in part by the **Corporation for National Research Initiatives,** using funds from the **Advanced Research Projects Agency** of the **United states Government** under Grant: **MDA972-92-J1029.**

The scanning agent for this project was the **Document Services** department of the **M.I.T Libraries.** Technical support for this project was also provided by the **M.I.T. Laboratory for Computer Sciences.**

Scanned

Date: 1/24/1995

M.I.T. Libraries
Document Services