MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 1218

August, 1991

Perceptual Organization, Figure-Ground, Attention And Saliency:

Figure Has A Fuzzy Boundary, Its Outside/Near/Top/Incoming Regions Are More Salient (Or Not; Or What?), And Holes Are Independent Of The Whole

J. Brian Subirana-Vilanova and Whitman Richards

Abstract: Figure and ground are often viewed as binary complements to one another, with a well defined boundary between them. A simple experiment shows otherwise: if the contour of a simple convex shape is perturbed to create a distinctive texture, it is typically the outside of the contour that provides the basis for similarity judgement, not the inside. The introduction of the appropriate task, however, can make the inside part of the contour become more salient. A similar result occurs for concave shapes, such as a C, where notions of "inside" and "outside" are not well defined. Here, as well as with "holes", any proposal that directly relates figure to fixed aspects of objects fails. This leads us to propose an operational definition of "figure".

Measures that assess similarity between shapes using a distance metric, cannot explain the above results. This leads us to suggest that there is a task-dependent bias in visual perception according to which the saliency of the two sides of a contour (inside and outside) is not the same. We suggest novel related biases such as "near is more salient than far", "top is more salient than bottom"and "expansion is more salient than contraction". We also discuss implications to visual perception; our findings seem to indicate that a frame is set in the image prior to recognition, and agree with a model in which recognition proceeds by the successive processing of convex chunks of image structures defined by this frame.

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. WR was also supported under NSF-IRI8900267.

1 Introduction

The natural world is usually conceived as being composed of different objects such as chairs, dogs or trees. This conception carries with it a notion that objects occupy a region of space, and have an "inside". By default, things outside this region of space are considered "outside" the object. Thus, the lungs of a dog are inside the dog, but the chair occupies a different region and is outside the object dog. When an object is projected into an image, these simple notions lead to what appears to be a clear disjunction between what is considered figure, and what is ground. Customarily, the figure is seen as the inside of the imaged shape as defined by its bounding contours (i.e. its silhouette). The region outside this boundary is ground. This implies that the points of an image are either figure or ground. Such a view is reinforced by reversible figures, such as Rubin's vase-face or Escher's patterns of birds and fish. Here, we show that such a simple disjunctive notion of figure and ground is incorrect, and that in general, the assignment of "figure" to the region of an image is ill-posed and consists of a fuzzy boundary.

If figure has an ill-defined boundary, then are there some regions of the image that are receiving "more attention" than others? We contend that the answer is yes and that this is due not only to processing constraints but also to some computational needs of the perceiver. In particular, a fuzzy boundary leaves room to address regions of the image that are of immediate concern, such as the handle of a mug (its outside) that we are trying to grasp, or the inside surface of a hole that we are trying to penetrate.

The ambiguity in defining a precise region of the image as figure arises in part because many objects in the world do not have clearly defined boundaries. Although objects occupy a region of space, the inside and outside regions of this space are uncertain. For example, what is the inside of a fir tree? Does it include the region between the branches where birds might nest, or the air space between the needles? If we attempt to be quite literal, then perhaps only the solid parts define the tree's exterior. But clearly such a definition is not consistent with our conceptual view of the fir tree which includes roughly everything within its convex hull. Just like the simple donut, we really have at least two and perhaps more conceptualizations of inside and outside. For the donut, the hole is inside it, in one sense, whereas the dough is inside it in another. But the region occupied by the donut for the most part includes both. Similarly for the fir tree, or for the air space of the mouth of a dog when it barks. Which of these two quite distinct inclusions of inside should be associated with the notion of figure?

In this paper we will present some suggestions that attempt to answer these issues. Since they are coupled to some fundamental problems of visual perception such as perceptual organization, attention, reference frames and recognition, it will be necessary to address these, too. The suggestions are based on some simple observa-

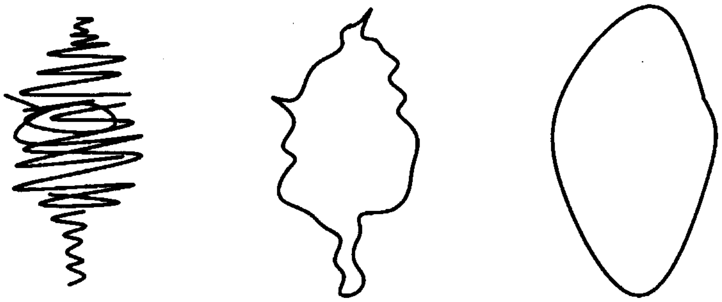

Figure 1: **Fir tree at several scales of resolution. What is inside the tree?**

tions. The essence of them can be easily grasped by the reader by glancing at the figures of the paper. The text alternates the presentation of such observations with the discussion of the suggested implications.

We begin, in the next section, by presenting some demonstrations that clarify how figural assignments are given to image regions. Along the way, we use these demonstrations to suggest a new, operational definition of "figure". In the following two sections we suggest that outside is more salient than inside; or not; or what? In section 5 we review the notion of "hole" and in sections 6 and 7 that of figure. In sections 8 and 9 we discuss the implications of our findings to visual perception. In section 10, we suggest that typically "near is more salient than far" and point to other similar biases. We end in section 11 with a summary of what is new about this paper.

2 Fuzzy Boundaries

Typically, figure-ground assignments are disjunctive, as in the Escher-drawings. However, when the image of a fractal-like object is considered, the exact boundary of the image shape is unclear, and depends upon the scale used to analyze the image. For the finest scale, perhaps the finest details are explicit, such as the needles of a spruce or the small holes through which a visual ray can pass unobstructed. But at the coarsest scale, most fractal objects including trees will appear as a smooth, solid convex shape. Any definition of figure must address this scale issue. Consider then, the following definition of figure:

Definition 1 (Figure): *"Figure" is that collection of image structures which currently are supporting the analysis of a scene.*

By this definition, we mean to imply that the perceiver is trying to build or recover the description of an object (or scene) in the world, and his information-processing

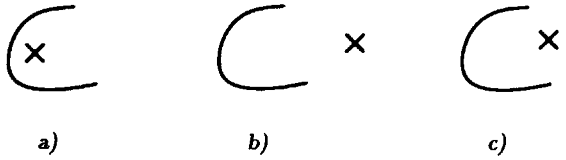

Figure 2: The notion of "what is figure" does not require that the figure be a region enclosed by a visible contour. In (a) the x is seen to lie within the C, and is associated with the figure, whereas in (b) the x lies outside the figure. In (c) the answer is unclear.

capability is focused on certain regions in the scene that are directly relevant to this task. The precise regions of the scene that are being analyzed, and their level of detail will be set by the demands of the goal in mind. Such a definition implies that the regions of the scene assigned as figure may not have a well-defined, visible contour; and leads to the following claim:

Claim 1 *The region of the image currently assigned as "figure" may not have a well-defined boundary earmarked by a visible image contour.*

In support of this claim, and hence our initial definition of "Figure", consider the C of Figure 2. Although the image contour by itself is well-defined, the region enclosed by the C is not. We would like to argue that the region "enclosed" by the "C" is a legitimate figural assertion. For example, if one asks the question does the "X" lie inside the C, our immediate answer is yes for case (a), and no for case (b). To make this judgement, the visual system must evaluate the size of the interior region of the C. Thus, by our definition, the concept "inside of C" must lead to an assignment of certain pixels of the display as figure. Without an explicit contour in the image, however, where should one draw the boundary between the figure, and its complement, the ground? For example, should we choose to close the figure with a straight line between the two endpoints? Another possibility would be to find a spline that completes the curve in such a way that the tangent at the two endpoints of the C is continuous for the complete figure. (Oddly, most observers choose neither, but rather something more closely approaching a "blurred" version of the C, as if they were using a large Gaussian mask on a colored closed C.) We contend that such "fuzzy" figural boundaries occur not only within regions that are incompletely specified, such as that within the incomplete closing of the C, but also within regions that appear more properly defined by explicit image contours.

To further clarify our definition of "figure", note that it is not prescribed by the retinal image, but rather by the collection of image structures in view. Any pixel-based definition of figure tied exclusively to the retinal image is inadequate, for it will not allow figural assertions to be made by a sequence of fixations of the object.

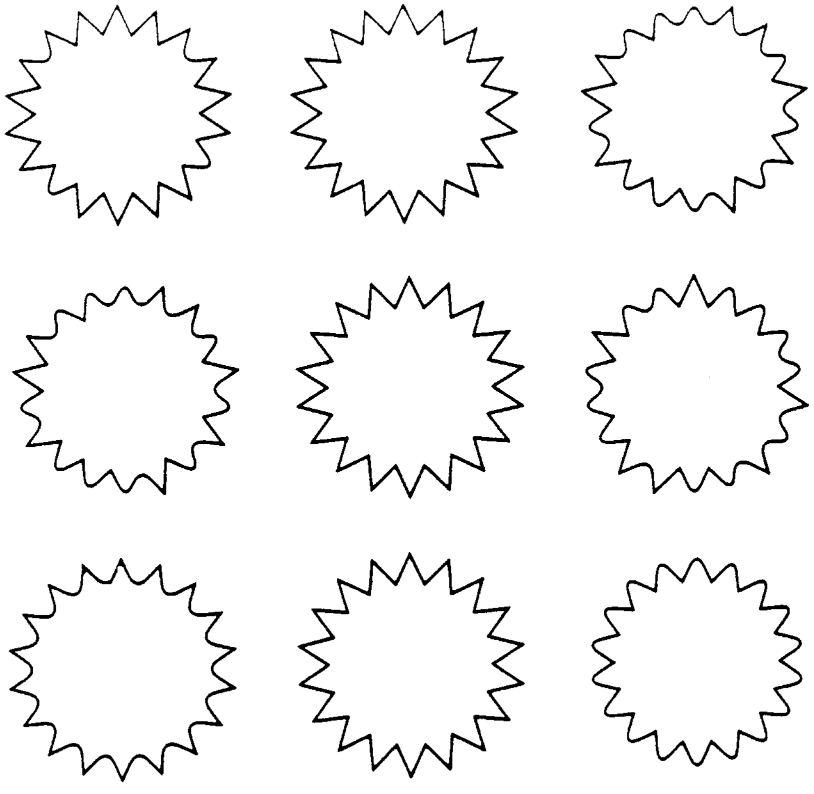

Figure 3: *Top row:* **Use the middle pattern as reference. Most see the left pattern as more similar to the reference. This could be because it has a smaller number of modified corners (with respect to the center) than the right one, and therefore, a pictorial match is better.** ***Second row:*** **In this case, the left and right stars look equally similar to the center one. This seems natural if we consider that both have a similar number of corners smoothed.** ***Third row:*** **Most see the left pattern as more similar despite the fact that both, left and right, have the same number of smoothed corners with respect to the center star. Therefore, in order to explain these observations, one cannot base an argument on just the number of smoothed corners. The position of the smoothed corners need be taken into account, i.e. preferences are not based on just pictorial matches. Rather, here the convexities on the outside of the patterns seem to drive our similarity judgement.**

Rather, a structure-based definition of figure presumes that the observer is building a description of an object or event, perhaps by recovering object properties. The support required to build these object properties is what we define as figure. This support corresponds closely to Ullman's incremental representations [Ullman 1984] upon which visual routines may act, and consequently the operations involved in figural assertions should include such procedures as indexing the sub-regions, marking these regions, and the setting of a coordinate frame. We continue with some simple observations that bear on these problems.

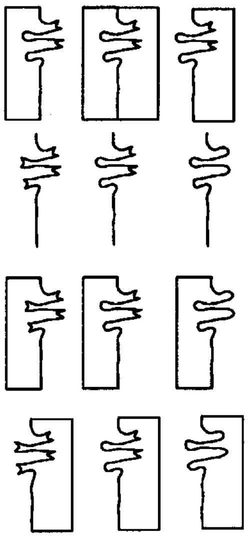

Figure 4: *Top:* Reversible figure. *Second Row:* The contour (shown again in the center) that defined the previous reversible figure is modified in two similar ways (left and right contours). *Third and fourth row:* When such three contours are closed a preference exists, and this preference depends for most on the side used to close the contour. Use the center shape as reference in both rows. As in the example of the previous Figure most favor the outer portions of the shape to judge similarity. A distance metric, based solely on a pictorial match and that does not take into account the relative location of the different points of the shape, cannot account for these observations.

3 Outside is More Salient than Inside

When binary figure-ground assignments are made for an image shape with a well-defined, simple, closed contour, such as an "O", the assignment is equivalent to partitioning the image into two regions, one lying inside the contour, the other outside. For such a simple shape as the "O", the immediate intuition is that it is the inside of the contour which is given the figural assignment, and this does not include any of the outside of the contour (see [Hoffman and Richards 1984] for example, where shape descriptors depend on such a distinction). By our definition, however, "figure" might also include at the very least a small band or ribbon outside the contour, simply because contour analysis demands such. As a step toward testing this notion, namely that a ribbon along the outer boundary of the shape should also be included when figural assignments are made, we perturb the contour to create simple textures such a those illustrated in Figure 3 (bottom row).

In this figure, let the middle star-pattern be your reference. Given this reference pattern, which of the two adjacent patterns is the most similar? Virtually everybody we test immediately pick the pattern on the left ($N > 20$)[1]. Now look more closely at these two adjacent patterns. In the left pattern, the intrusions have been smoothed, whereas in the right pattern the protrusions are smooth. Clearly the similarity judgement is based upon the similarity of the protrusions, which are viewed as sharp convex angles. The inner discrepancy is almost neglected.

The same conclusion is reached even when the contour has a more part-based flavor, rather than being a contour texture, as in Figure 4. Here, a rectangle has been modified to have only two protrusions[2]. Again, subjects will base their similarity judgments on the shape of the convex portion of the protrusion, rather than the inner concavity.

This result is not surprising if shape recognition is to make any use of the fact that most objects in nature can be decomposed into parts. The use of such a property should indeed place more emphasis upon the outer portions of the object silhouette, because it is here that the character of a part is generally determined, not by the nature of its attachment. Almost all attachments lead to concavities, such as when a stick is thrust into a marshmallow. Trying to classify a three-dimensional object by its attachments is usually misguided, not only because many different parts can have similar attachments, but also because the precise form of the attachments is not reliably visible in the image. Hence the indexing of parts (or textures) for shape recognition can proceed more effectively by concentrating on the outer extremities.

Another possible justification for such observations is that, in tasks such as grasping or collision avoidance, the outer part is also more important and deserves more attention because it is the one that we are likely to encounter first[3].

The outer region of a shape is thus more salient than its inner region. This implies that the region of scene pixels assigned to figure places more weight on the outer, convex portions of the contour than on its interior concave elements (or to interior homogeneous regions), and leads to the following claim:

Claim 2 *The human visual system assigns a non-binary figure-ground function to scene pixels with greater weight given to regions near the outside of shapes, which become more salient.*

[1] The results are virtually independent of the viewing conditions. However, if the stars sustain an angle larger than 10 degrees, the preferences may reverse.

[2] This is one sample from a collection of similar shapes that was used in an experiment too informal to report, but which clearly corroborated the observations presented in this paper.

[3] For example, in Figure 3 (bottom), the center and left stars would "hurt" when grasped, whereas the right star would not because it has a "smooth" outside.

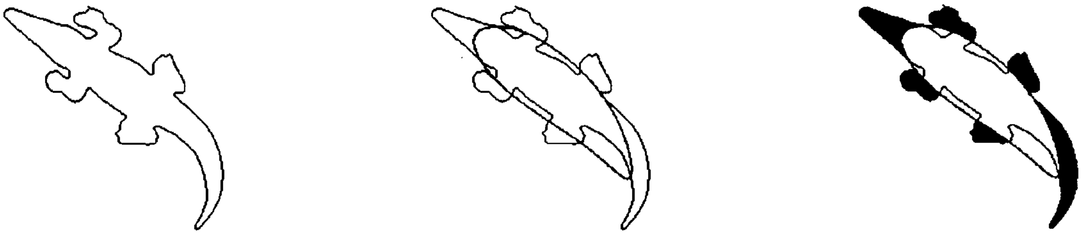

Figure 5: *Left:* Alligator. *Center:* Alligator with *frame curve* superimposed. The frame curve has been computed using a standard smoothing algorithm. *Right:* The outside of the different alligator's parts has been shaded by a simple coloring operation. The frame curve and the silhouette of the shape have been used as bounding contours for the coloring operation. Note that there is one shaded region per part.

Note that this claim simply refers to "regions near the outside", not to whether the region is convex or concave. In Figure 4, the outer portion of the protrusion contains a small concavity, which presumably is the basis for the figural comparison.

Exactly what region of the contour is involved in this judgement is unclear, and may depend upon the property being assessed. All we wish to claim at this point is that whatever this property, its principal region of support is the outer portion of the contour. The process of specifying just which image elements constitute this outer contour is still not clear, nor is the measure (nor weight) to be applied to these elements. One possibility is an *insideness* measure. Such a measure could be easily computed as a function of the distance of the image elements to the "smoothed" version of the contour (a circle in Figure 3 and something close to a rectangle in Figure 4). In this context, the smoothed contour corresponds to the notion of frame curves as used in [Subirana-Vilanova 1991].

This leads us to the following definition of frame curve which has to be read bearing in mind claim 1:

Definition 2 (Frame Curve): *A frame curve is a virtual curve in the image which lies in "the center" of the figure's boundary.*

In general, the frame curve can be computed by smoothing the silhouette of the shape. This is not always a well-defined process because the silhouette may be ill-defined or fragmented, and because there is no known way of determining a unique scale at which to apply the smoothing. Figure 5 (center) shows the frame curve for an alligator computed using such scheme. On the right, the regions of the shape that are "outside" the frame curve have been colored; note that these regions do not intersect, and correspond closely to the outer portions of the different parts of the shape. As mentioned above, these outer portions are both more stable and more likely to be of immediate interest.

Note that Claim 2 supports Claim 1 because the figure-ground function mentioned in Claim 2 is to be taken to represent a fuzzy boundary for figure. The notion of frame curve should not be seen as a discrete boundary for figure (perhaps only at a first approximation). Indeed, we contend that a discrete boundary is not a realistic concept.

4 Inside is More Salient than Outside

Consider once more the three-star patterns of Figure 3. Imagine now that each of these patterns is expanded to occupy 20 degrees of visual angle (roughly your hand at 30 centimeters distance). In this case the inner protrusions may become more prominent and now the left pattern may be more similar to the middle reference pattern. (A similar effect can be obtained if one imagines trying to look through these patterns, as if in preparation for reaching an object through a hole or window.) Is this reversal of saliency simply due to a change in image size, or does the notion of a "hole" carry with it a special weighting function for figural assignments?

For example, perhaps by viewing the central region of any of the patterns of Figure 3 as a "hole", the specification of what is outside the contour has been reversed. Claim 2 would then continue to hold. However, now we require that pixel assignments to "figure" be gated by a higher level cognitive operator which decides whether an image region should be regarded as an object "hole" or not.

5 When a Hole Is Not a Hole

Consider next the star patterns in Figure 6, which consist of two superimposed convex shapes, one inside the other. Again with the middle pattern as reference, most will pick as most similar the adjacent pattern to the right. This is surprising, because these patterns are generally regarded as textured donuts, with the inner-most region a hole. But if this is the case and our previous claim is to hold, then the left pattern should have been most similar. The favored choice is thus as if the inner star pattern were viewed as one object occluding another. Indeed, if we now force ourselves to take this view, ignoring the outer pattern, then the right patterns are again more similar as in Figure 3. So in either case, regardless of whether we view the combination as a donut with a hole, or as one shape occluding part of another, we still use the same portion of the inner contour to make our similarity judgement. The hole of the donut thus does not act like a hole. The only exception is when we explicitly try to put our hand through this donut hole. Then the inner-most protrusions become more salient as previously described for "holes".

These results lead to the following claim:

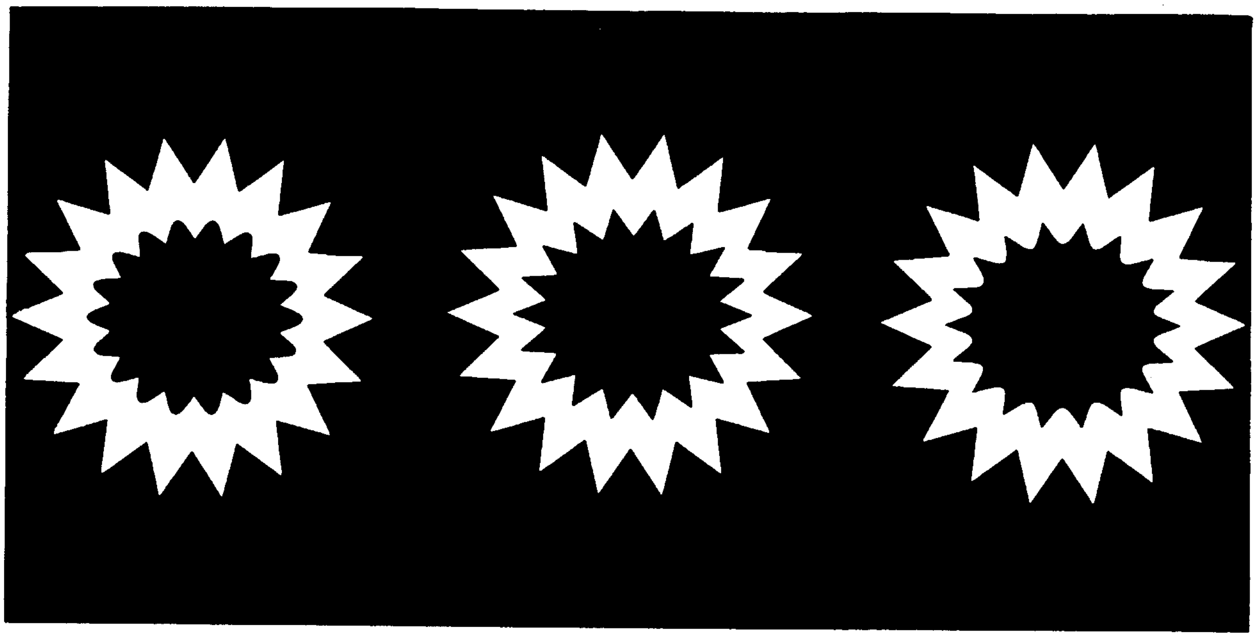

Figure 6: Star patterns with "holes" treat the inside ring of the shape as if this ring was an occluding shape, i.e. as if it was independent from the surrounding contours, even if one perceives a donut like shape.

Claim 2 (revisited): *Once the attentional frame is chosen, then (conceptually) a sign is given to radial vectors converging or diverging from the center of this frame (i.e. the focal point). If the vector is directed outward (as if an object representation is accessed), then the outer portion of the encountered contours are salient. If the vector is directed inward to the focal point (as if a passageway is explored), then the inner portion of the contour becomes salient*[4]

So far we have been ignoring one important step in the processing of visual information: attention. It has long been known that humans concentrate the processing of images in certain regions or structures of the visual array. Attention has several forms: one of them, perhaps the most obvious, is gaze. We cannot explore a stationary scene by swinging our eyes past it in continuous movements. Instead, the eyes jump with a saccadic movement, come to rest momentarily and then jump to a new locus of interest (see [Yarbus 1967]).

[4] There is an interesting exception to the rule: if the sise of the contours is very big (in retinal terms) then the inside is always more salient (as if we where only interested in the inside of large objects).

In the star patterns that we discussed in Section 2 (see Figure 3) the attention was focused primarily on the stars as whole objects. That is, there is a center of the figure that appears as the "natural" place to begin to direct our attention. The default location of this center, which is to become the center of a local coordinate frame, seems to be, roughly, the center of gravity of the figure [Richards and Kaufman 1969], [Kaufman and Richards 1969], [Palmer 1983]. Attention is then allowed to be directed to locations within this frame. Consider next the shapes shown in the top of Figure 7. Each ribbon-like shape has one clear center on which we first focus our attention. So now let us bend each of the ribbons to create a new frame center which lies near the inner left edge of each figure (Figure 7, lower). Whereas before most regarded the left pattern as more similar to the middle reference, now the situation is starting to become confused. When the ribbons are finally closed to create the donuts of figure 6, the favored similarity judgement is for the right pattern. The primary effect of bending and closing the ribbon seems to be a shift in the relation between the attentional frame and the contours. Following the center of gravity rule, this center eventually will move outside the original body of the ribbon. This suggests that the judgments of texture similarity are dependent on the location of the attentional coordinate frame.

Typically, as we move our gaze around the scene, the center of the coordinate frame will shift with respect to the imaged contours, thus altering the pixel assignments. A shift in the focus of attention without image movements can create an effect similar to altered gaze. More details regarding these proposed computations will be given in the next sections. What is important at the moment is that the saliency of figural assignments will depend upon the position of the contour with respect to the location of the center of the attentional coordinate frame. The reader can test this effect himself by forcing his attention to lie either within or outside the boundary of the ribbons. Depending upon the position chosen, the similarity judgments change consistently with Claim 2.

As we move our attention around the figure, the focus of attention will shift, but the frame need not. But if the frame moves, then so consequently will the assignment of scene pixels to (potentially) active image pixels. The visual system first picks a (virtual) focal point in the scene, typically bounded by contours, and based on this focal point, defines the extent of the region (containing the focal point) to be included as "figure". If all events in the selected region are treated as one object or a collection of superimposed objects, then the radially distant (convex) portions of the contours drive the similarity judgments and are weighted more heavily in the figural computations. On the other hand, if the choice is made to regard the focal point as a visual ray along which something must pass through (such as a judgement regarding the size of a hole), then the contours that lie radially the closest are given greater weight (i.e. those that were previously concave). This led us to the revised version of claim 2, namely that the attentional coordinate frame has associated with it either

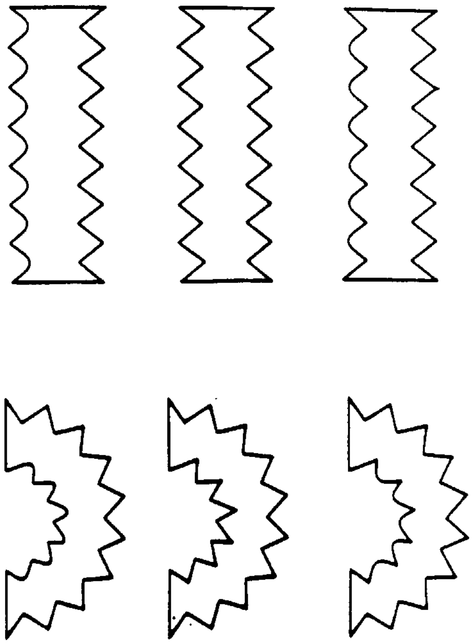

Figure 7: *Top:* **Using the middle shape as a reference, most see the left shape as more similar.** *Bottom:* **If this same shape is bent, the situation becomes confused.**

an inward or outward pointing vector that dictates which portion of a contour will be salient (i.e. outer versus inner). We have argued that the orientation of this vector is task-dependent.

Here we must introduce a contrary note. We do not propose that the attentional frame is imposed only upon a 3D structure seen as an object. Such a view would require that object recognition (or a 2 1/2 sketch) had already taken place. Rather, our claim is that the attentional coordinate frame is imposed upon a (frontal plane) silhouette or region prior to recognition, and is used to support the object recognition process, such as by indexing the image elements to a model. Hence, because a commitment is made to a coordinate frame and the sign of its object-associated vectors (inward or outward), proper object recognition could be blocked if either the location of the frame or the sign of its radial vectors were chosen improperly. When 3D structure is obtained and analogous 3D frames are imposed, similar claims are possible but they are not the subject of this section.

6 Against Ground

Our latest claim introduces a complementary, mutually exclusive state to the attentional coordinate frame within which figural processing is integrated. When the attentional vector is pointing outward, as in "object mode", does this then imply that the contour regions associated with the inward state of this vector should be assigned to a separate state called "ground"? We argue against such an assignment, especially the view that objects in the foreground yield figure and that the rest of the display is ground.

Consider the following experiment of [Rock and Sigman 1973] in which they showed a dot moving up and down behind a slit or opening, as if a sinusoidal curve was being translated behind it. The experiments were performed with slits of different shapes, so that in some cases the slit was perceived as an occluded surface and in others as an occluding one. They found that the perception of the curve is achieved only if the slit is perceived as an occluded region and not when it is perceived as an occluding region. Using their terms, the "correct" perception is achieved only if the slit is part of ground but not when it is part of the figure. Instead, we suggest that what is figure has not changed but rather its *attributes* did because the slit was viewed as a passageway between objects, and not as an object with a hole. This is an entirely different concept of "ground".

In support of our view, another experiment by [Rock and Gilchrist 1975] shows that figure need not correspond to the occluding surface. In this second experiment, they showed a horizontal line moving up and down with one end remaining in contact with one side of an outline figure of a face. Consequently, the line in the display changes in length. When the line is on the side of the face most observers see it changing size, adapting to the outline, while when it is on the other side of the contour, it is seen with constant length but occluded by the face. This has been described as a situation in which no figure-ground reversal occurs. However, in our terms the figure has changed because the attended region changes. In the first case figure corresponds to the occluding surface, and in the second to the occluded one. Thus, figure need not correspond to the occluding surface, even when the surfaces that are occluded are known. Again, this conclusion is consistent with our definition of figure and our claims.

7 What's Figure?

Our least controversial claim is that the image region taken as figure depends upon one's attention and goal. Reversible illusory patterns, such as the Escher drawings or Rubin's face-vase support this claim. The more controversial claim is that the image region taken as figure does not have a boundary that can be defined solely in terms of an image contour, even if we include virtual contours such as those cognitive edges formed by the Kanizsa figures. The reason is two fold: First, the focal position of our attentional coordinate frame with respect to the contours determines that part of the contour used in figural similarity judgments, implying that the region assigned to figure has changed, or at the very least has been given altered weights. Second, whether the focal position is viewed as part of a passageway or alternatively simply as a hole in an object affects the figural boundary. In each case, the region is understood to lie within an object, but the chosen task affects the details of the region being processed. This effect is also seen clearly in textured C-shaped patterns, and becomes acute when one is asked to judge whether X lies inside the C, or if Y will fit into the C, etc. The virtual boundary assigned to close the C when making such judgments of necessity will also depend in part upon the size of the second object, Y. To simply assert that figure is that region lying inside an image contour misses the point of what the visual information processor is up to.

Again, a simple experiment from the Rock laboratory demonstrates that "figure" is not simply a region of the display, but rather a collection of scene elements. Some of the elements within this "attended" region may be ignored, and thus, not be part of the structures at which higher level visual operations are currently being applied. [Rock and Gutman 1981] showed two overlapping novel outline figures, one red and one green, for a brief period, e.g. one second. Subjects were instructed to rate figures of a given color on the basis of how much they liked them (this attracts attention to one of the figures). They later presented subjects with a new set of outline figures and asked subjects whether they had seen these figures in the previous phase of the experiment, regardless of their color. They found that subjects were very good at remembering the attended shapes but failed on the unattended ones. This experiment agrees with the model presented in this paper, in which only the attended set of structures is being processed. Thus, figure is, clearly, not a region of pixels contained in the attended figure because the attended figure was partly contained in such a region and did not yield any high-level perception.

In order to show that what they were studying was failure of perception and not merely selective memory for what was being attended or not attended, [Rock and Gutman 1981] did another experiment. They presented a series, just like in the previous case but with two familiar figures in different pairs of the series, one in the attended color and one in the unattended color. They found that the attended familiar figure was readily recognized but that the unattended familiar figure was not. It is natural that if the unattended figure is perceived and recognized it would stand out. Failure of recognition therefore supports the belief that the fundamental deficit is of perception. The extent of such deficit is unclear; it may be that the level of processing reached for the unattended figures is not complete but goes beyond that of figures not contained in the attended region.

Therefore, an operational definition of "what is figure?" seems more fruitful in trying to understand how images are interpreted. Our definition is in this spirit, and leads to a slightly different view of the initial steps involved in the processing of visual images than those now in vogue in computational vision. This is the subject of the next two sections.

8 Perceptual Organization, Object Types and Figural Assignments: Which Comes First and the Role of Convexity

There have been many proposals on what are the different steps involved in visual perception and it is not the main goal of this paper to make yet another such proposal. Nevertheless, our findings have some relevant implications to what should be the nature of these steps which we will now discuss.

We suggest that figure and objects are not strongly coupled. Figure is simply the image-based structures which support some high-level processing, regardless of whether the region is assumed to be an object in the foreground, an occluded object, a hole, a passageway *or none of the above.* This does not mean that the current assumption of which type of region is figure does not play any role in the processing. Rather, we have shown several examples where the assumptions or role of the region is transformed onto an attribute of figure that governs both which portions of the contours are included in the processing and the type of processing to be done in it. Curiously, this suggests that a cognitive judgement proceeds and selects that portion of an image or contour to be processed for the task at hand.

But how can a cognitive judgement anticipate where attention will be directed without some preliminary image processing that notes the current contours and edges? We are thus required to postulate an earlier, more reflexive mechanism that directs the eye, and hence the principal focus of attention, to various regions of the image. Computational studies suggest that the location of such focus may involve a bottom-up process such as the one described in [Subirana-Vilanova 1990]. Subirana-Vilanova's scheme computes points upon which further processing is directed using either image contours or image intensities directly. Regions corresponding to each potential point can also be obtained using bottom-up computations[5]. There is other computational evidence that bottom-up grouping and perceptual organization processes can correctly identify candidate interesting structures (see [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-

[5] The result of these computations may affect strongly the choice of reference frames. For example, if the inner stars in Figure 6 are rotated so as to align with the outer stars (creating convexities in the space between the two), our attention seems more likely to shift to the region in-between the two stars and in this case the similarities will change in agreement with claim 2.

Another way of increasing the preference for the "in-between" reference frame in Figures 6 and 7 is by coloring the donut black and leaving the surrounding white (because in human perception there is a bias towards dark objects).

Vilanova 1990]).

Psychological results in line with the Gestalt tradition [Wertheimer 1923], [Koffka 1935], [Köhler 1940] argue for bottom-up processes too. However, they also provide evidence that top-down processing is involved. Other experiments argue in this direction, such as the one performed by [Kundel and Nodine 1983] in which a poor copy of a shape is difficult, if not impossible to segment correctly unless one is given some high level help such as "this image contains an object of this type". With the hint, perceptual organization and recognition proceed effortlessly. Other examples of top-down processing include [Newhall 54], [Rock 1983], [Cavanagh 1991], C.M. Mooney and P.B. Porter's binary faces, and R.C. Jones' spotted dog. The role of top-down processing may be really simple, such as controlling or tweaking the behavior of an otherwise purely bottom-up process; or perhaps it involves selecting an appropriate model or structure among several ones computed bottom-up; or perhaps just indexing. In either case the role of top-down processing cannot be ignored. Indeed, here we claim that the setting up of the attentional coordinate frame is an important early step in image interpretation.

Our observations suggest that perceptual organization results in regions that are closed or convex (at a coarse scale) as discussed (see also section 5). This corroborates computational studies on perceptual organization which also point in that direction, demonstrating the effectiveness and the viability of perceptual organization schemes which limit themselves to finding convex or "enclosed" regions (or at least favor them) [Jacobs 1989], [Huttenlocher and Wayner 1990], [Subirana-Vilanova 1990], [Clemens 1991], [Subirana-Vilanova and Sung]. It is still unclear if this is a general limitation of the visual system, a compound effect with inside and outside, or rather specific to shape perception. There are, however, several areas that may bring some more light onto the question. One of them is the study of the gamma effect: When a visual object is abruptly presented on a homogeneous background, its sudden appearance is accompanied by an expansion of the object. Similarly, a contraction movement is perceived if the object suddenly disappears from the visual field. Such movements were observed a long time ago and were named "gamma" movements by [Kenkel 1913], (see [Kanizsa 1979] for an introduction). For non-elongated shapes, the direction of movement of the figure is generally centrifugal (from the center outward for expansion and from the periphery toward the center for contraction). For elongated shapes, the movement occurs mainly along the perceptual privileged axes. It is unclear whether the movements are involved in the selection of figure or if, on the contrary are subsequent to it. In any case they might be related to a coloring process (perhaps responsible for the expansion movements) involved in figure selection that would determine a non-discrete boundary upon which saliency judgments are established (see also [Mumford, Kosslyn, Hillger and Herrnstein 1987]). If this is true, studying the effect on non-convex shapes (such as those on Figure 7) may provide cues to what sort of computation is used when the figures are not convex, and to the nature of the inside/outside asymmetry.

Another area that may be interesting to study is motion capture which was observed informally by [Ramachandran and Anstis 83]: When an empty shape is moved in a dynamic image of random dots it "captures" the points that are inside it. This means that the points inside the shape are perceived as moving in the same direction of the shape even though they are, in fact, stationary (randomly appearing for a short interval). This can be informally verified by the reader by drawing a circle on a transparency and sliding it through the screen of a connected TV with noise: The points inside the circle will be perceived as moving along with the circle. The results hold even if the circle has some gaps and it has been shown that they also hold when the shapes are defined by subjective contours [Ramachandran 86]. There is no clear study of what happens for non-convex shapes such as a C. What portions are captured? Informal experiments done in our laboratory seem to confirm that the boundary of the captured region is somewhat fuzzy for unclosed shapes like a C which supports the notion of a fuzzy boundary. In addition, the shape for the captured region seems to have convexity restrictions similar to the ones suggested for the inside-outside relations. It is unclear if both mechanisms are related but the similarity is intriguing. This seems a very promising direction for future research.

Further evidence for the bias towards convex structures is provided by an astonishing result obtained recently by [Cumming, Hurlbert, Johnson and Parker 1991]: when a textured cycle of a sine wave in depth (the upper half convex, the lower half concave) is seen rotating both halfs may appear convex[6], despite the fact that this challenges rigidity[7] (in fact, a narrow band between the two ribbons is seen as moving non-rigidly!).

It is also of interest to study how people perceive ambiguous patterns or tilings [Tuijl 1980], [Shimaya and Yoroizawa 1990] that can be organized in several different ways. It has been shown that in some cases the preference for convex structures can overcome the preference for symmetric structures that are convex [Kanizsa and Gerbino 1976]. The interaction between convex and concave regions is still unclear, especially if the tilings are not complete.

Studies with pigeons[8] [Herrnstein, Vaughan, Mumford and Kosslyn 1989] indicate that they can deal with inside-outside relations so long as the objects are convex but not when they are concave. It is unclear if some sort of "inside-outside" is used at all by the pigeons. More detailed studies could reveal the computation involved, and

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

[7] This observation will be relevant later because it supports the notion that a frame is set in the image before structure from motion is recovered (see claim 3 and related discussion).

[8] The pigeon visual system, despite its reduced dimensions and simplicity, is capable of some remarkable recognition tasks that do not involve explicit inside/outside relations. See [Herrnstein and Loveland 1964], [Cerella 1982], [Herrnstein 1984] for an introduction.

perhaps whether they use a local feature strategy or a global one. This, in turn, may provide some insights into the limitations of our visual system.

9 What is the Shape of Reference Frames?

As described in the previous sections, our proposal implies that the establishment of a frame of reference is required prior to recognition. In other words, without the frame, which is used to set the saliency of the different image regions, recognition cannot proceed. We have pinned down three aspects of it: its location, its size and its inside and outside. Previous research on frames has focused on the 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], [Schwartz 1981], [Shepard and Cooper 1982], [Jolicoeur and Landau 1984], [Jolicoeur 1985], [Palmer 1985], [Corballis and Cullen 86], [Maki 1986], [Jolicoeur, Snow and Murray 1987], [Parsons and Shimojo 1987], [Robertson, Palmer and Gomez 1987], [Shepard and Metzler 1988], [Corballis 1988], [Palmer, Simone and Kube 1988], [Georgopoulos, Lurito, Petrides, Schwartz and Massey 1989], [Tarr and Pinker 1989]), on the influence of the environment ([Mach 1914], [Attneave 1968], [Palmer 1980], [Palmer and Bucher 1981], [Humphreys 1983], [Palmer 1989]), on its location ([Richards and Kaufman 1969], [Kaufman and Richards 1969], [Cavanagh 1978], [Palmer 1983], [Cavanagh 1985], [Nazir and O'Reagan 1990]), and on its size ([Sekuler and Nash 1972], [Cavanagh 1978], [Jolicoeur and Besner 1987], [Jolicoeur 1987], [Larsen and Bundsen 1987]). Exciting results have been obtained in this directions but it is not the purpose of this paper to review them.

The shape of the frame, instead, has received very little attention. [Subirana-Vilanova 1990] proposed that in some cases, a curved frame might be useful (see also [Palmer 1989]). In particular, he proposed to recognize elongated curved objects, such as the ones shown in Figure 7, by unbending them using their main curved axis as a frame to match the unbended versions. If human vision used such a scheme, one would expect no differences in the perception of the shapes shown on the top of Figure 7 from those on the bottom of the same figure. As we have discussed, our findings suggest otherwise, which argues against such a mechanism in human vision.

10 Related Effects: What Do You Want to be More Salient?

The shapes used so far in our examples have been defined by image contours. The results, however, do not seem to depend on how such contours are established and similar results seem to hold when the shapes are defined by motion or other discontinuities. Thus, the results seem to reflect the true nature of shape perception. In this section we will suggest that similar biases in saliency occur in other dimensions of visual perception. What all of them have in common is that they require the

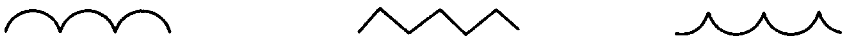

Figure 8: Top is more salient than bottom: Using the middle pattern as reference, most see the right contour as more similar.

establishment of an attentional frame of reference at a very early stage, and that the nature of the frame depends on the task at hand. In particular, we will suggest that: top is more salient than bottom, near is more salient than far and outward motion is more salient than inward motion.

Top is more salient than bottom; or not.

Consider the contours in Figure 8, the center contour appears more similar to the one on the right than to the one on the left. We suggest that this is because the top of the contours is, in general, more salient than its bottom. We can provide functional justification similar to that given in the inside-outside case: the top is more salient because, by default, the visual system is more interested in it, as if it were the part of a surface that we contact first. Just like with our inside-outside notion, the outcome can be reversed by changing the task (consider they are the roof of a small room that you are about to enter). Thus, there is an asymmetry on the saliency of the two sides of such contour (top and bottom) similar to the inside/outside one discussed in the previous sections.

Near is more salient than far; or not.

When looking for a fruit tree of a certain species it is likely that, in addition we are interested in finding the one that is closer to us. Similarly, if we are trying to grasp something that is surrounded by other objects, the regions that are closer to our hand are likely to be of more interest than the rest of the scene. We suggest that when three-dimensional information is available, the visual system emphasizes the closer regions of the scene. Evidence is shown in Figure 9 in which we show a stereo pair with surfaces similar to the silhouette of the star of Figure 3.

At a first glance, most see two of the three surfaces of Figure 3 as being more similar. The preference, as in the previous case, can be reversed if we change the task: imagine, for example, that you are flying above such surfaces and are looking for a place to land. Your attention will change to the far portions of the surfaces and with it your preferred similarities. Therefore, attention and the task at hand play an important role in determining how we perceive the three-dimensional world. Note also, that, as in the previous examples, a matching measure based on the distance between two surfaces cannot account for our observations. For in this case, such

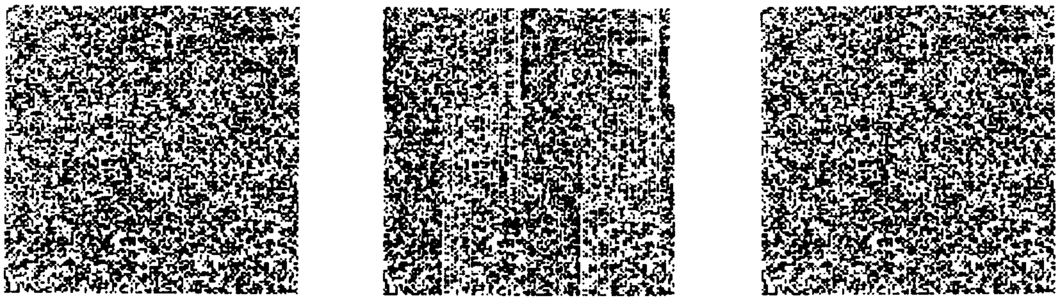

Figure 9: This random dot stereo diagram illustrates that close structures are more salient in visual perception than those that are further away.

distance to the center surface is the same for both bounding surfaces.

Expansion is more salient than contraction; or not.

Is there a certain type of motion that should be of most interest to the human visual system? Presumably, motion coming directly toward the observer is more relevant than motion away from it. Or, similarly, expanding motion should be more salient than contracting motion. Evidence in support of this suggestion is provided by a simple experiment illustrated in Figure 10[9]. Like in the previous cases, two seemingly symmetric percepts are not perceived equally by the visual system. This distinction, again, seems to bear on some simple task-related objectives of the observer.

So, what's more salient? How does perception work?

Inside/outside, near/far, expansion/contraction and top/bottom are generally not correlated. If saliency were determined independently for each of these relations, then conflicts could arise in some cases. For example, the inside of an object may be near or far, in the top or in the bottom of the image. Will, in this case, the outside regions on the bottom be more salient than those that are inside and on the top?

This is an important issue that will not be addressed in this paper. A more detailed understanding of how attention and perceptual organization interact with the early vision modules is required. In any case, it would be interesting to find a modular division showing how these processes may interact. Unfortunately, this is a no-win situation. Either the modules are too few to be interesting or the division

[9] In a pool of 7 MIT graduate students, all but one reported that their attention was directed first at the expanding pattern.

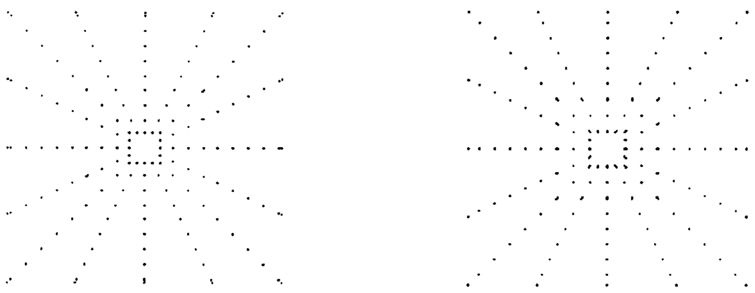

Figure 10: When the dots in this two similar figures are moving towards the center on one of them, and towards the outside on the other, attention focuses first on the expanding flow. This provides evidence that motion coming directly toward the observer is more salient than motion away from it.

is easily proven to be wrong. Nevertheless, it may be useful to give a proposal as precise as possible to illustrate what has been said so far. Figure 11 is it.

Like in [Witkin and Tenenbaum 1983], our proposal is that grouping is done very early (before any 2 1/2 D sketch-like processing), but we point out the importance of selecting a coordinate frame which, among other things, is involved in top-down processing and can be used to index into a class of models. Indexing can be based on the coarse description of the shape that the frame can produce, or on the image features associated with the frame. As shown in Figure 11, this frame may later be enhanced by 3D information coming from the different early vision modules. Like in [Jepson and Richards 91], we suggest that one of the most important roles of the frame is to select and articulate the processing on the "relevant" structures of the image (see also footnote 7). This leads us to the last claim of this paper:

Claim 3 *An attentional "coordinate" frame is imposed in the image prior to constructing an object description for recognition.*

11 What's New

The fact that figure and ground reversals are attention related has been known for some time [Rubin 1921][10]. However, there appears to be no precise statement of the relation between "figure" and notions of "inside" and "object", nor has it been

[10] [Rubin 1921] showed subjects simple contours where there was a two way ambiguity in what should be figure (similar to the reversible figure in the top of Figure 4). He found that if one region was found as figure when shown the image for the first time then, if on subsequent presentations the opposite region was found as figure, recognition would not occur.

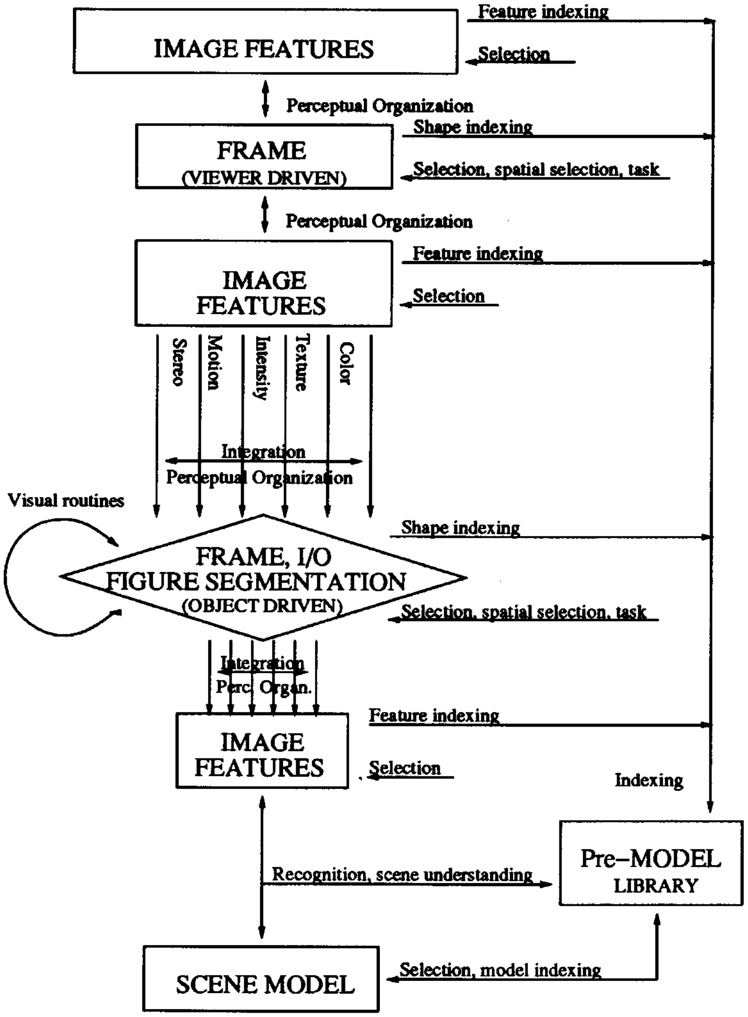

Figure 11: A modular description of visual perception that illustrates some of the concepts discussed in this paper. The different frames and image features depicted may share data structures; this accounts for some implicit feed-back in the diagram. (The Figure emphasizes the order in which events take place, not the detailed nature of the data structures involved.) It is suggested that perception begins by some simple image features which are used to compute a frame that is used to interpret these image features. The frame is an active structure which can be modified by visual routines. In this diagram, shape appears as the main source for recognition but indexing plays also an important role. Indexing can be based on features and on a rough description of the selected frame.

noted previously that contour saliency depends on *inside/outside, near/far, expansion/contraction and top/bottom* relations, and changes when the *task* is changed, such as viewing a region as something to pass thru, rather than as a shape to be recognized.

These new observations support an operational definition of *figure* which is based on attention and which rules ground out of the picture. We have suggested that occlusion be treated as an attribute of figure. An alternative way of saying the same, is to call figure (as defined in this paper) *the processing focus* and to associate a figure/ground attribute to it. Regardless of the syntaxis, any proposal that relates figure to object fails: A hole can become figure while the inside of a shape can become "ground". The idea that figure has a non-discrete boundary has not been suggested previously either. This leads to the concept of *frame curve* which can be used for *shape segmentation* in conjunction with Inside/Outside relationships.

Our findings also demonstrate that the task at hand controls *top-down processing*. Existing evidence for top-down processing shows its role in increasing the speed and performance of recognition (by providing hints, such as restricting the set of models to be considered). However, a qualitative role of top-down processing (such as determining whether we are looking for an object or a hole), not dependent on the image, like the one presented here has not been suggested previously.

We have shown that "matching to model" will not correspond with human perception unless inside/outside, top/bottom, expansion/contraction and near/far relations are factored early in the recognition strategy. We have also discussed several ways in which the role of *convexity* can be studied in human vision, such as inside/outside relations, gamma movements and motion capture. Our observations provide new light into the nature of the attention and perceptual organization processes involved in visual perception. In particular, they indicate that a *frame* is set in the image prior to recognition and agree with a model in which recognition proceeds by the successive processing of *convex* chunks of *image structures* defined by this frame.

Acknowledgments

Part of this work was done while JBSV was at the NTT Human Interface Laboratories in Yokosuka (Japan) during the winter of 1990; thanks to Kosugi-san, Kaneda-san and Sato-san for interesting discussions. Thanks also to Shimon Ullman, Ellen Hildreth, David Beymer, Tomaso Poggio and Eric Grimson for interesting discussions and suggestions for improving the presentation of this paper.

References

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

[2] 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.

[3] E.G. Boring. *A history of experimental psychology.* Appleton Century Crofts, Inc., 1964. Second Edition. Originally published 1929.

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

[5] 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.

[6] P. Cavanagh. What's up in top-down processing? In Andrei Gorea, editor, *Proceedings of the XIIIth european conference on visual perception.* 1990.

[7] J. Cerella. Mechanisms of concept formation in the pigeon. In D.J. Ingle, M.A. Goodale, and R.J.W. Mansfield, editors, *Analysis of visual behavior*, pages 241–260. The MIT Press, Cambridge and London, 1982.

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

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

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

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

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

[13] 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.

[14] R.L. Gregory. *The intelligent eye.* McGraw-Hill Book Company, New York, St. Louis and San Francisco, 1970.

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

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

[17] R.J. Herrnstein. Objects, categories, and discriminative stimulus. In H.L. Roitblat, T.G. Bever, and H.S. Terrace, editors, *Animal Cognition: Proceedings of the Frank Guggenhelm Conference.* Lawrence Erlbaum Associates, Hillsdale N.J, 1984.

[18] R.J. Herrnstein, W. Vaughan Jr., D.B. Mumford, and S.M. Kosslyn. Teaching pigeons an abstract relational rule: Insideness. *Percepion and Psychophysics*, 46(1):56–64, 1989.

[19] R.J. Herrnstein and D. Loveland. Complex visual concept in the pigeon. *Science*, 46:549–551, 1964.

[20] 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.

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

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

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

[24] R.C. James. The dalmatian dog, photographed by R.C. James has been reproduced in a number of different publications. A binary version can be found in [Marr 82] and a non-discretized one in [Gregory 1970].

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

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

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

[28] 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.

[29] 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.

[30] 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.

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

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

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

[34] F. Kenkel. Untersuchungen uber den zusammenhang zwischen erscheinungsgrosse und erscheinungsbewegung bei einigen sogennanten optischen tauschungen. *Zeitschrift fur Psychologie*, 67:358–449, 1913.

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

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

[37] H.L. Kundel and C.F. Nodine. A visual concept shapes image perception. *Radiology*, 146(2):363–368, 1983.

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

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

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

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

[42] 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.

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

[44] D. Marr. Analysis of occluding contour. *Proceedings of the Royal Society of London B*, 197:441–475, 1977.

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

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

[47] D. Mumford, S.M. Kosslyn, L.A. Hillger, and R.J. Herrnstein. Discriminating figure from ground: The role of edge detection and region growing. *Proceedings of the National Academy of Science*, 87:7354–7358, 1984.

[48] 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.

[49] S.M. Newhall. Hidden cow puzzle picture. *American Journal of Psychology*, 65(110), 1954.

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

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

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

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

[54] 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.

[55] 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.

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

[57] 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.

[58] P.B. Porter. Another picture puzzle. *American Journal of Psychology*, 67:550–551, 1954.

[59] V.S. Ramachandran. Capture of stereopsis and apparent motion by illusory contours. *Percepion and Psychophysics*, 39(5):361–373, 1986.

[60] V.S. Ramachandran and S.M. Anstis. Displacement thresholds for coherent apparent motion in random dot-patterns. *Vision Research*, 23(12):1719–1724, 1983.

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

[62] 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.

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

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

[65] I. Rock and A. Gilchrist. The conditions for the perception of the covering and uncovering of a line. *American Journal of Psychology*, 88:571–582, 1975.

[66] I. Rock and D. Gutman. The effect of inattention on form perception. *Journal of Experimental Psychology: Human Perception and Performance*, 7:275–285, 1983.

[67] I. Rock and E. Sigman. Intelligence factors in the perception of form through a moving slit. *Perception*, 2:357–369, 1973.

[68] E. Rubin. *Visuell Wahrgenommene Figuren*. Glydendalske, 1921. See [Boring 1964] Pg. 605, or [Rock 83] Pg. 306.

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

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

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

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

[73] 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.

[74] A. Shimaya and I. Yoroizawa. Automatic creation of reasonable interpretations for complex line figures. In *Proceedings Int. Conf. on Pattern Recognition*, pages 480–484, Atlantic City, New Jersey, 1990.

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

[76] J.B. Subirana-Vilanova. The skeleton sketch: Finding salient frames of reference. In *Proceedings Image Understanding Workshop, 1990*, pages 399–414. Morgan and Kaufman, 1990. Also in the proceedings of ICCV'90.

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

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

[79] J.B. Subirana-Vilanova and K.K. Sung. In preparation.

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

[81] A. Treisman and G. Gelade. A feature integration theory of attention. *Cognitive Psychology*, 12:97–136, 1980.

[82] H. Tuijil. Perceptual interpretation of complex line patterns. *Journal of Experimental Psychology: Human Perception and Performance*, 6(2), 1983.

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

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

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

[86] 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.

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

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

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
August 1991

3. REPORT TYPE AND DATES COVERED
memorandum

4. TITLE AND SUBTITLE
Perceptual Organization, Figure-Ground, Attention and Saliency

5. FUNDING NUMBERS
S1-801534-2
DACA76-85-C-0010
N00014-85-K-0124
NSF-IRI8900267

6. AUTHOR(S)
J. Brian Subirana

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Artificial Intelligence Laboratory
545 Technology Square
Cambridge, Massachusetts 02139

8. PERFORMING ORGANIZATION REPORT NUMBER
AD-A259964

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
Information Systems
Arlington, Virginia 22217

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AIM 1218

11. SUPPLEMENTARY NOTES
None

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution of this document is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Abstract: Figure and ground are often viewed as binary complements to one another, with a well defined boundary between them. A simple experiment shows otherwise: if the contour of a simple convex shape is perturbed to create a distinctive texture, it is typically the outside of the contour that provides the basis for similarity judgement, not the inside. The introduction of the appropriate task, however, can make the inside part of the contour become more salient. A similar result occurs for concave shapes, such as a C, where notions of "inside" and "outside" are not well defined. Here, as well as with "holes", any proposal that directly relates figure to fixed aspects of objects fails. This leads us to propose an operational definition of "figure".

Measures that assess similarity between shapes using a distance metric, cannot explain the above results. This leads us to suggest that there is a task-dependent bias in visual perception according to which the saliency of the two sides of a contour (inside and outside) is not the same.

(continued on back)

14. SUBJECT TERMS (key words)
vision

15. NUMBER OF PAGES
29

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED

Block 13 continued:

We suggest novel related biases such as "near is more salient than far", "top is more salient than bottom"and "expansion is more salient than contraction". We also discuss implications to visual perception; our findings seem to indicate that a frame is set in the image prior to recognition, and agree with a model in which recognition proceeds by the successive processing of convex chunks of image structures defined by this frame.

CS-TR Scanning Project
Document Control Form

Date : 11 / 03 /94

Report # Aim - 1218

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: 30
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 Form2(PGS □ Funding Agent Form □ Cover Page
□ Spine □ Printers Notes □ Photo negatives
□ Other: ________________

Page Data:

Blank Pages(by page number):________________

Photographs/Tonal Material (by page number):________________

Other (note description/page number):
Description : Page Number:

Scanning Agent Signoff:
Date Received: 11 / 03 / 94 Date Scanned: 11 / 04 / 94 Date Returned: 11 / 10 / 94

Scanning Agent Signature: Michael W. Cook