maps are independent for each site, and parallel algorithms for line following are known and can easily be adapted. Also, calculations are simple and stable, as no curvatures or any other derivatives need to be computed on the digital curves.

The system can rank features based on their perceptual importance. This allows a real-time application to process as many features as time permits.

Some of the issues which have not been addressed are the resolution dependency of the description. At this time, only one level of description is possible. Also, we have not tried to localize end-points of curves ending abruptly. Since all computations are performed on a discrete grid, quantization and rounding errors restrict the selectivity and amount of clutter the system can handle.

Clearly, more work is necessary to improve the performance of the end-point field. Currently, only very simple examples have been tested, and no effort has been made to integrate responses from the various fields. The most straightforward way to combine responses would be to add (point-to-point) the co-variance matrices of the Extension, Point, and End-point fields, followed by an explicit edge extraction step (e.g. marching lines). A more intelligent approach would try to resolve conflicts between representations produced by the three fields separately.
account the interference by computing the eccentricity measure at each site. Table 2 summarizes the main differences and similarities of the two methods.

**Table 2: Comparison between the Hough Transform and Our Scheme**

<table>
<thead>
<tr>
<th></th>
<th>Our scheme</th>
<th>Hough Transform</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Complexity</strong></td>
<td>Same for all shapes and properties</td>
<td>Grows with the number of parameters</td>
</tr>
<tr>
<td><strong>Space Requirements</strong></td>
<td>Constant (O(\text{image size}))</td>
<td>Grows with the number of parameters and resolution.</td>
</tr>
<tr>
<td><strong>Properties (Families of shapes) coding</strong></td>
<td>Yes, by defining suitable fields</td>
<td>Not in any obvious way</td>
</tr>
<tr>
<td><strong>Analytical shapes</strong></td>
<td>Yes, by the same method</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Voting pattern</strong>(^a)</td>
<td>2-D</td>
<td>1-D(^b)</td>
</tr>
<tr>
<td><strong>Localization of shape</strong></td>
<td>Yes</td>
<td>Not always. Further processing is needed to find location</td>
</tr>
<tr>
<td><strong>Means of output</strong></td>
<td>A ridge of maxima</td>
<td>A peak in parameter space</td>
</tr>
<tr>
<td><strong>Choice of bin size</strong></td>
<td>Always image resolution</td>
<td>Hard (May be crucial)</td>
</tr>
<tr>
<td><strong>Selectivity</strong></td>
<td>Low (Because of compaction)</td>
<td>High</td>
</tr>
<tr>
<td><strong>Transformation</strong></td>
<td>Non-Linear</td>
<td>Linear</td>
</tr>
</tbody>
</table>

\(^a\) Our nomenclature. Refer to text for definition.

\(^b\) Or at least one dimension less that of the problem.

### 3.12 Conclusion and Future Work

We have introduced a unified way to extract perceptual features in edge images. By ‘unified’ we mean that all low-level features (edgels, points) are treated in a uniform way, and no special cases exist. The scheme is threshold-free and non-iterative. It is especially suitable for parallel implementation, since computations of the saliency...
have in the desired fashion. Also, note that if our voting method is kept, we can detect junctions at the same time.

To implement a circle finder using our scheme, we would use our Extension Field, but with all field strengths set to the value 1. Coding of the other constraints\(^1\) is where the strengths of the individual field elements come into play. *This is not possible with the original Hough Transform.*

The downside is that the result is not found at an isolated peak of the parameter space, but rather as a continuous ridge of peaks (in the case of the Extension Field). Also, note that the parameter field we use is *always* the image plane, which, in many cases, is smaller than a classical Hough parameter space. This ‘Compaction’ of data is the cause for the low selectivity compared to Hough transforms.

However, for Perceptual Grouping, this loss of selectivity seems to be an advantage. Note that the Hough Transform could find clusters (e.g. straight lines) even when they are not salient to humans, because the notion of interfering features is not made explicit. This makes the Hough Transform a *linear superposition* which is generally not suitable for perceptual grouping. Our scheme, on the other hand, is highly non-linear and takes into

\(^1\) that make it a perceptually appealing scheme
cessing is needed to localize the actual segment in the image plane. Note that this process is absolutely global, as it ignores distances between contributing candidates. If some orientation data is known at a site, fewer candidates gather votes, and the ability to reveal the desired shape gets better.

The same scheme can be extended for other shapes [4] by extending the parameter space.

### 3.11.2 Our scheme as a Hough transform

Our system can be viewed as a Hough Transform where the parameter space is the image plane itself. This allows for many more degrees of freedom in the choice of shapes, and in the basic definition of the desired shapes. This choice of the parameter space allows us to define other voting patterns\(^{12}\) which enable us to encode the constraints (see 3.5.1).

Our scheme is thus capable of finding shapes described by their properties (smoothness etc.) rather than by their exact analytical parameters (as in the Hough Transform).

Note that the classical Hough transform for edgels can be implemented in our scheme. We replace the Extension Field by a straight line, and remove the attenuation factors. For non-oriented inputs, the Point field (again without the attenuation) will be-

\(^{12}\) Similar to the sinusoid defined for the straight line detector
other. In some cases, responses from the two fields should enhance, rather than inhibit, each other. Heitger and van der Heydt [26] show ways to integrate what they call the Para and Ortho fields. They use a look-up-table to determine locally what the combined response should be. Williams’ [82] technique can also be used to disambiguate this type of responses by imposing a 3-D occluding surface constraint. With this constraint, it is likely that a floating disk occluding a set of lines will be the prevailing explanation.

3.11 Comparison With the Hough Transform

3.11.1 The classical Hough transform

Consider he classical Hough transform [18,30] which can detect co-linear clusters in a dot image in the presence of noise. This is accomplished by detecting peaks in an accumulator array. If the line is parameterized in terms of \((\theta, d)\), each point votes in \(\theta, d\) space for a sinusoid with equal weights. The votes in each cell are scalar, and the accumulation is simply a sum. This scheme is trivially extended to oriented edges, in which case, each edgel votes for a cell.

For more complex shapes, the scheme can be generalized by increasing the dimensionality of the accumulator space. For instance, circle detection requires 3 dimensions (2 for the center, 1 for the radius).

The next phase in the process is to search for peaks in the parameter space array. The desired feature is only defined in terms of the above parameters and further pro-
3.10.4 Experimenting with the End-Point field

We tested the end-point field with the synthetic image in Figure 3.34. It is clear that the outer circle receives relatively low saliency, and can be completely removed if we consider single votes in the plane as wrong hypotheses (see Figure 3.38).

3.10.5 End-point and Extension field interaction

In many cases, the response from the Extension field and that from the End-point field will conflict. A simple example is Figure 3.38(a). Here, the Extension field will attempt to connect lines across the middle circle, while the End-point offers the circular interpretation. We currently do not have a mechanism to inhibit one response or the other.
Many other experiments (such as in [7]) indicate that end-point formations that require concave T-junctions are not normally perceived. Also, real objects tend to have longer convex boundaries, than concave ones (see [28]), and since boundary length is proportional to the probability of an intersection, convex T-junctions are more common.

### 3.10.3 Building the End-Point Field

The angle distribution derived previously suggests convolving our original Extension field with a multi-directional edge having a diameter function of a sinusoid, as was done for the point field, and shown in Figure 3.37. This kind of field will vote strongly for straight angle T-junctions, and weakly for other directions. This merely means that more support is needed for more acute angles.
The probability of that angle being between 60 and 90 is thus 0.5.

Adding more lines to the system will not change the distribution of angles, since we can consider all segments to be pair-wise independent.

The same result holds for the 3-D case, since the distribution of the projection (either orthographic or perspective) of any line to 2-D is equal to the marginal distribution of the original 3-D distribution. That is, uniformly distributed segments in space project to uniformly distributed segments in the plane.

Piecewise smooth curved objects can be approximated by linear segments to any degree of accuracy, and should thus satisfy the same distribution. The same is true for lines of different lengths. It is easy to see that the length of the segment does not change the probability of a given intersection angle.

### 3.10.2 Convex T-junctions are more common

We believe that the a priori probability of having convex T-junctions is larger than concave T-junctions. This observation can be illustrated through a simple example, as shown in Figure 3.36.

Figure 3.38(a) is another example where the inner circle is perceived, but the other one is not. Again, we claim this is because concave T-junctions are not normally hypothesized.
look at the distribution of the intersection angles between the two segments. We claim that angles close to 90 degrees are more likely to appear than acute intersection angles. (We do not count cases where the segments do not intersect.)

Proof: Without loss of generality we assume that the first segment is centered at the origin of the plane along the x-axis. Since all positions in the plane have equal probability, the probability of intersection is proportional to the area of the locus of positions allowing intersection for a certain angle. This is a parallelepiped, as shown in Figure 3.35. The area of this parallelepiped is proportional to $\sin \alpha$ for $\alpha$ between 0 and 90 degrees. We can thus write the distribution function of intersection angles as:

\[
F(\alpha) = P(x \leq \alpha) = \begin{cases} 
0, & \alpha < 0 \\
\alpha, & \alpha \geq 0 \leq 90 \\
1, & \alpha > 90 
\end{cases}
\]

Clearly, angles close to 90 degrees are more probable. □
plemented by Heitger and von der Heydt [26], in the case where end-points are already marked. Illusory edges appear to outline the said occluding shape. Figure 3.34(a) illustrates an end-point scenario.

![Figure 3.34](image)

**Figure 3.34** An end-point formation. (a) A center egg-like shape is not only perceived but also looks whiter (after [7]). (b) An invisible circle occludes lines. No sensation of a circle is evident, because angles of intersection are not suitable. (c) The inner circle is perceived, but the outer one is not!

### 3.10.1 Straight angles in T-Junctions are more likely than any other

We derive the distribution of T-junction angles in a random world and show that close-to-90° T-junction angles are more likely to appear than any other angle. We claim that our human perception attempts to perform perceptual tasks on end-points stimuli only when the illusory intersection angles justify it. In Figure 3.34(a) all lines meet the illusory contour with almost straight angles, thus making the shape ‘visible’. In Figure 3.34(b) the lines are occluded by an exact circle, but the angles are much more acute. *No* perception of shape is evident, even though the end-points trace an exact circle.

**Claim**: Given two unit size segments, we independently drop each one of them on a finite board, with uniform probability with respect to position and angle. We then...
3.9.4 Complexity Issues

A naive way to implement the algorithm requires $O(n^2k)$ operations, where $n$ is the side size of the image, and $k$ is the number of edge elements in the input image. This complexity assumes that the Extension Field’s size is $n^2$, and we need to position it over every element of the input. In practice, the local density of edgels restricts the useful scope of the field. This means that a smaller finite field can be used. The complexity becomes now $O(Ck)$, where $C$ is the number of non-zero elements of the field. This last modification has the disadvantage of not being able to bridge gaps larger than the size of the field. Alternatively, instead of computing a dense saliency map, we can compute the saliency of existing edgels only. Here we assume the existing edgels are stored in some condensed array, so that traversal of the whole image is not necessary. Instead, every edgel votes for all others with the appropriate Extension field entry. This results in complexity of $O(k^2)$, and can be useful as a focus of attention map. This mode allows then for a second pass on the salient features only by employing some pre-defined threshold. A parallel implementation can bring the full algorithm to $O(C)$.

3.10 Application of scheme to end points (End-point field)

In many cases we tend to interpret end-points as being a partially occluded line. If enough end-points are available, they support the hypothesis of a shape occluding a collection of lines. This is a well documented phenomenon, and has been partially im-
Figure 3.33  (a)-(e) extracted curves for 2, 4, 6, 8 and 10% of additive noise. (f) input with 10% of noise
ample, may find line formations even in a very cluttered image, where such formations are not visible, and could be a result of accidental alignments.

We have performed a controlled experiment to determine the amount of random noise which still allows for a correct following of a curve, given a constant factor of existing edge. The test is whether all points along the perceptual circle belong to the local maxima, as defined before. Only uniform noise in space and orientation was applied to the original image. The circle has a radius of 25 pixels and consists of 33 segments of unit strength and the percentages of noise applied are 2, 4, 6, 8, and 10 percent\textsuperscript{11}, again of unit strength. This means that about a quarter of the circle edge exists. The results of marching along all ridges are shown in Figure 3.33. Up to 6 percent of noise, the circle (or large parts of it) is recoverable. With 8 percent noise and more, the saliency map degrades, and the marching lines algorithm is only able to extract partial arcs of the original circle. The original image with 10\% of noise is also given in Figure 3.33, as reference.

\textsuperscript{11} that is, each pixel has a 0.1 probability of becoming an erroneous site.
3.9.3.5 Noise Breakdown Point

Here we show that out algorithm degrades gracefully, and does not attempt to create formations when they are not perceptually salient. The Hough transform, for ex-
3.9.3.3 Point Field

We tested our system on the image in figure 3.31(a). Initially, the system was run using the Point field. This resulted in a saliency map with orientation data. A second phase of computation was then performed, using the directional Extension field (Figure 3.6). That stage produced the final saliency map as shown in figure 3.31(b).

3.9.3.4 Straight line Field

We tested our straight line operator on the previous example, as shown in Figure 3.31. Clearly the straight line appears to be salient. The ellipse is still vaguely visible, since it can be viewed as being piecewise straight to some degree.
example of the steps involved in producing a high-level description of a given image, using the junction map in conjunction with the saliency map. The result of following the strongest ridges for the Kanizsa square are shown in Figure 3.30. Note that the circles have a much lower saliency, due to their relatively high curvature.
Figure 3.29(b) shows that lower saliency curves also exist in the final result, and a higher-level process needs to deal with them. These curves, however, have normally a much lower saliency as can be seen from the 3-D plot in Figure 3.29(c). Figure 3.27 also shows the result of the same procedure for the other scenes. Figure 3.28 shows an
the array, as well as the accuracy of the implementation. Thus, this phenomenon is not a limitation of the theory, but rather a limitation of the discrete implementation.

3.9.3.2 Synthetic Images

We have tested our approach with the synthetic data shown earlier in Figure 1.2. The saliency map produced is shown (strength only) as a grey-level image in figure 3.27. The result of following the strongest crest-line are shown in Figure 3.29(a).
votes for more than one curve. Such multiple voting is not inhibited in any way, and no decision is made regarding which vote is correct. Moreover, the system is designed to report all possible formations, without checking whether there is any physical evidence in the input. Encoding such constraints into the existing methodology may be possible, e.g. as a second pass over the votes accumulator.

The last basic arrangement involves a scenario where two parallel lines are a small distance apart. By bringing the lines closer and closer together it is possible to evaluate the degree of discriminability of the system. The actual experiment takes two converging lines and attempts to march along the resulting ridges. Figure 3.26(a) shows the input, and Figure 3.26(b) shows the result of marching along the highest ridges. It appears that when the lines are about 3 sites apart they merge into a single line. Note that the ability to keep two lines separate is a function of the resolution of
Figure 3.24  
(a) an input image with edgels separated by increasing number of pixels.  
(b) a plot of the saliency profile.  
Note that the first 3 pairs from the top give rise to a single-peak profile, while the rest ‘sag’ in the middle.

the circles, but also have high agreement in the area between them.  
This artifact appears because the theory fails to handle explicitly the situation where a single edgel
3.9.3 Experimental Results

3.9.3.1 Basic arrangements

Before demonstrating results on complex inputs, we show here how the system behaves when simple arrangements of few edgels exist in the input.

The first case is a co-linearity test involving exactly two co-linear edgels at a given distance apart. It is interesting to look at the profile of the saliency as a function of the distance between the two.

As our theory suggests, the straight line connecting two co-linear edgels is always along a ridge of saliency (i.e. more salient than its environment), but at the same time decays with distance to the closest edgel. When the two edgels are very close (as shown in Figure 3.24) the saliency profile between them can in fact be larger than their own saliency. This does not conflict with the theory, but merely demonstrates the fact that a dense smooth curve is likely to have a ‘center’ more salient than its edges. In Figure 3.24(a) we generated pairs of edgels with increasing separation. Figure 3.24(b) shows a plot of the computed saliency. The ridges are clearly seen on all but the bottom-most pair. At some distance apart the middle part of the imaginary connecting line starts ‘sagging’.

Our next example demonstrates the ghosting effect that occurs when certain symmetrical configurations vote for locations in space which have no physical meaning. Take, for example, Figure 3.25. here two fragmented circles vote correctly along
with the theory, since high agreement values yield higher votes, which in turn produce
larger standard deviation. Determining the threshold would thus involve finding the
value which is a pre-defined multiplier of the standard deviation in the scene.

Another way of estimating true saliency is to refrain from normalization during
the process of computing the saliency map. It is then possible to compare the maxi-
mum saliency (as discussed in section 3.7.4) to the actual saliencies in the scene.
of all vectors against an arbitrary vector, and flipping the vector if the sign is negative.

In pseudo-code:

```plaintext
if (dot_product(vector1,vector2)<0) flip(vector2);
if (dot_product(vector1,vector3)<0) flip(vector3);
if (dot_product(vector1,vector4)<0) flip(vector3);
```

where flip is

\[
flip \left( \begin{bmatrix} v_x, v_y \end{bmatrix} \right) = \begin{bmatrix} -v_x, -v_y \end{bmatrix}
\]

It is now possible to impose additional constraints on the extracted curves, such as minimum length, hysteresis thresholds, and maximum curvature. Imposing these (and possibly other) constraints is left to higher-level processes.

### 3.9.2 Determining absolute saliency and noise threshold

Determining the strength level which can no longer be considered salient may prove important in certain situations. Recall that all maps represent relative likelihood values, and any given absolute saliency value has no meaning outside the context of the complete map. In fact, even a nonsense input would give rise to high saliency values in various random locations in the image.

In trying to estimate the level of “useful” saliency we looked at the histogram of votes for well structured inputs compared to nonsensical inputs, in the presence of noise. Figure 3.23 plots the histogram of two sample inputs. It appears that the standard deviation of ‘real’ data is larger than that of a random set. This is in accordance
where two curves go through the same site. It is possible to resolve the ambiguity in the
direction of the edges by looking at the bigger picture and imposing some simple
continuity constraints. This has not been done in our work, mainly because such cases
are rare.

The actual location of the intersection of the edge with the sides of the site ($p$ in
Figure 3.22) is determined through a first-order approximation using the computed
values at the vertices. This simply amounts to:

$$p = \frac{\text{Value of Vertex 1}}{\text{Value of Vertex 1} - \text{Value of Vertex 2}}$$  \hspace{1cm} (3.15)

It is also possible to employ a second (or higher) order approximation (using fur-
ther away value) to get a better approximation. This has not been implemented,
though.

**Alignment**

An important issue in ensuring the consistency of the evaluation of the four ver-
tices is the alignment of the normal vectors. Recall that the front-end of the system
(the one that produces the saliency map) cannot ensure that all normals (or tangents)
belonging to an underlying curve are oriented consistently along the curve. Thus, the
Marching Squares algorithm must locally align the four normal vectors before apply-
ing equation (3.13) to them. The alignment involves testing the sign of the dot product
Equation 3.13 assigns the value and sign of the gradient in the direction of the normal to the scalar field. Now every site is examined for zero-crossing, and a process called a “Marching lines” algorithm extracts final polygonal line description of the scene. In essence, the marching process starts by identifying a beginning of a curve. In our system this is done by looking in the vicinity of the site with the highest saliency. This step yields a small line segment. We then follow one of the endings of this segment a ‘march’ to the site pointed by that ending. The process is repeated until we either reach the boundary of the image, or arrive back at the starting point.

The “Marching squares” algorithm (after [41]) examines the sign of the four corners of every site in the image. Exactly sixteen combinations exist, out of which fourteen give rise to edges. Figure 3.22 depicts some common combinations, and their respective edges. Case (d) in Figure 3.22 is interesting because it describes situations where two solutions exist, and cannot be resolved locally.

Figure 3.22 Some typical scenarios encountered by the Marching Squares algorithm. (a) a curve crosses the site diagonally. (b) top to bottom. (c) no curve since all vertices are of the same sign. (d) an ambiguous case where two solutions exist, and cannot be resolved locally.
The reason is that the system is not sensitive enough to keep two (or more) close-by junctions separate, so there is no point to search for them in the saliency map. In our current implementation this separation distance is estimated at 5 pixels.

### 3.9.1.2 Extracting curves

To achieve a sub-pixel resolution when extracting curves, we look at the differential properties of the crest lines in the saliency map. Since tangents are given at every site, it is possible to convert this maximal map to a zero-crossing map. This is done by looking at the dot-product of the gradient and the normal vector to the curve at each point. This dot-product should be zero along the crest line, because the gradient there is exactly zero. It also changes signs when moving across a crest line, thus producing a zero-crossing map. The new zero-crossing (scalar) array \( S \) is thus computed as follows:

\[
S(i,j) = \tilde{g}_{i,j} \cdot \tilde{n}_{i,j}
\]  

(3.13)

where \( g \) is computed by

\[
\bar{g}_{i,j} = [\|SM(i+1,j)\| - (\|SM(i-1,j)\|, \|SM(i,j+1)\| - \|SM(j-1,i)\|)]
\]  

(3.14)

\( SM \) being the enhanced saliency map. Since derivatives are computed, a smoothing step is applied to the saliency map to avoid small local maxima which would interfere with the extraction of the major curves.
A multiple-resolution description then, is merely a description of a scene as viewed from several different distances. This is done by either convolving the input with multi-resolution fields, or by using a pyramid type of input.

Extracting features at different resolutions can be achieved by convolving the input with a set of fields, each dedicated to an interval of curvatures. The sum of all these fields equals the original Extension Field\textsuperscript{10}. This method decomposes the image into curvature scale space, and not the classic size scale space.

### 3.9 High-level Feature Extraction

#### 3.9.1 Description (Curves And Junctions)

Once a saliency map and a junction saliency map are acquired, a process that actually groups salient shapes is started.

#### 3.9.1.1 Extracting Junctions

Extracting junctions involves finding all local maximums and sorting them by value. The process can stop when the absolute value goes below a predefined noise level. We later discuss some methods to determine what the noise threshold is for a given saliency map. In practice we smooth the junction map before searching for maximum peaks, and also mask off a neighborhood around a detected peak. The smoothing operation is meant to eliminate small harmonies from becoming junctions, and the masking to avoid close-by peaks which are beyond the selectivity range of the system.

\textsuperscript{10} For the curve saliency map.
Knowledge of the location of junctions is crucial in the linking phase of any edge detection scheme. As we show in the results, linking is performed by following curves from one junction to the next.

\[ \lambda_{\text{min}} \text{ is large} \]

*Figure 3.21* Junction sites are formed when two or more sets of edgels vote for a single site, but from two (or more) distinct directions.

### 3.8 Multiple Resolution

We have shown that perceptual grouping is a global process, and as such requires that large portions of a scene be visible to the eye (or possibly to the fovea) at the any one time. It is not clear to us whether a sequential scan of a very large image can trigger a perception of a feature spanning the whole image. This makes the process highly sensitive to viewing distance, for two reasons:

- Perceived curvatures change with viewing distance.
- Features may fully appear at one resolution, but only partially at another.
3.7.5 Detection of Junctions

A junction is defined as a salient point having a low eccentricity value.

Regular (non-junction) points along a curve are expected to have high eccentricity values. On the other hand, junction points are expected to have low eccentricity, since votes were accumulated from several different directions. By combining the eccentricity and the eigenvalue at a point, we acquire a continuous measure of the likelihood of that site being a junction. We redefine our previous definition of eccentricity slightly, so that low eccentricity scores high, or:

\[(E = \frac{\lambda_{\text{min}}}{\lambda_{\text{max}}}) \Rightarrow 0 \leq E \leq 1\]  \hfill (3.11)

The product of our new eccentricity measure and the raw saliency measure \(\lambda_{\text{max}}\) yields the junction saliency operator:

\[E \cdot \lambda_{\text{max}} = (\frac{\lambda_{\text{min}}}{\lambda_{\text{max}}}) \cdot \lambda_{\text{max}} = \lambda_{\text{min}}\]  \hfill (3.12)

This process creates a Junction Saliency map. Interestingly enough, this map evaluates to just \(\lambda_{\text{min}}\) at every site (as shown in (3.12)), which simply means that the largest non-eccentric sites are good candidates for junctions. By finding all local maxima of the junction map we localize junctions (see results in Figure 3.28).

We are not able, however, to determine from the information at the site, what the angles of the voting curves are.
The low eccentricity in the vicinity of the correct curve is due to the large variance of votes in these areas. These votes along the correct curve, when accumulated, all contribute to sites, thus making the sites almost non-directional.

It may look at first sight that the field, while voting for the correct curve, contaminates the environment by voting for many other cells in the image. This can be regarded as noise, and is inherent to the process. However, while the fields voting for a curve agree along the curve, they disagree in any other area of the image. This means that the contribution of a complete curve to the environment is almost isotropic and cannot affect the direction of true votes in any area. It has been shown in the previous section that close-by areas get low agreement values, and far away areas get low voting weight. In both cases, the interference is small. This is true for all smooth curves in the image.

**Noise Contamination.** Random segments do contaminate the environment, and their effect is reduced through the robustness of the voting scheme. Clearly, high levels of erroneous edgels could interfere with the system’s ability to extract the correct descriptions. In the experimental results section we empirically test for allowable levels of noise.
3.7.4.2 The Non-maximum Suppression Phenomenon

We have mentioned the superior selectivity of the Enhanced Saliency map. To illustrate this behavior, we look at the eccentricity only map of a straight line (Figure 3.20(a)). Note how low the eccentricity is close to where the real line passes.

In Figure 3.20(b) we see the raw saliency map of the same line. The Enhanced Saliency map is simply the product of the two maps, point by point, and sharpens the edges of the correct curves, thus creating a non-maximum suppression effect (Figure 3.20(c)). Similar maps for a perfect circle are depicted in Figure 3.20(d-f).
3.7.4 Properties of the extension field

3.7.4.1 A longer line implies a stronger and more directed field, but up to a point.

Using a simple example, we demonstrate the behavior of the field when extending a straight line. Figure 3.19 shows a cross-section of a saliency map computed on a series of straight lines with increasing lengths. Clearly, the saliency grows as a function of the length, and the map becomes more directed (thinner ridge). Also, saliency converges to some finite value which is just the infinite integral along the main axis of the Extension Field. This can be explained intuitively by noting that the influence of a given edge segment decreases monotonically with distance, so a given site can only gather votes up to a certain “influence radius”. This observation can be used to estimate absolute saliency.

Figure 3.19 Saliency of a line does not grow forever. It converges to some value which is the infinite straight integral of the extension field.
a site far away from where the ‘action’ is, which accepts exactly one vote. (This can happen in practice.) The eccentricity value is 1, but the site is of no importance.

However, Consider $\lambda_{\text{max}}$ itself. We have,

$$\frac{\lambda_{\text{min}} + \lambda_{\text{max}}}{2} \leq \lambda_{\text{max}} \leq \lambda_{\text{min}} + \lambda_{\text{max}}$$  \hfill (3.9)

By (3.9) $\lambda_{\text{max}}$ is bounded from both sides by the proximity measure in (3.8) and has the eccentricity coded into it. When the value leans towards the left side of (3.9), eccentricity is low and vice-versa.

**Thus, $\lambda_{\text{max}}$ is chosen as the raw saliency measure in our scheme.**

This choice however, may still amplify locations which are very strong in terms of number of votes, but weak in eccentricity\(^9\). The product of E and $\lambda_{\text{max}}$ produces the desired result, termed the *enhanced saliency* measure $SM$, or:

$$SM = \lambda_{\text{max}} \cdot (1 - \lambda_{\text{min}} / \lambda_{\text{max}}) = \lambda_{\text{max}} - \lambda_{\text{min}}$$  \hfill (3.10)

**Thus, $\lambda_{\text{max}} - \lambda_{\text{min}}$ is chosen as the enhanced saliency measure.**

It is important to note that other functions of the eigenvalues can also satisfy the same conditions of monotonicity, but the ones chosen seem to be the *simplest* possible indicators of the desired behavior.

\(^9\) For example, accumulation points and junctions! (where $\lambda_{\text{min}} \equiv \lambda_{\text{max}}$)

Chapter 3-Our Approach (2-D) 58
Note that the angle $\theta$ has disappeared on the r.h.s. of (3.7). This means that the sum of eigenvalues is independent of the orientations of the voting vectors and can hence be used as an indicator of proximity (a wider sense of proximity of course), and as a naive saliency measure.

Equation (3.7) can be written as:

$$\lambda_{\min}^{t+1} + \lambda_{\max}^{t+1} = m_{20} + m_{02} + (R\cos\theta)^2 + (R\sin\theta)^2 = m_{20}^t + m_{02}^t + R^2 \quad (3.7)$$

Where $N$ is the number of segments in the original image.

We define the eccentricity $E = 1 - \lambda_{\min}/\lambda_{\max}$ as a measure of agreement. This value is between 0 and 1 (Since $\lambda_{\min} \leq \lambda_{\max}$ and are both non-negative). Our intuitive notion of ‘agreement’, or of a majority vote on a continuous scale, is consistent with the above definition. This means that in all cases where we feel that collection $A$ has better ‘agreement’ than collection $B$, the corresponding eccentricity values will share the same relationship (i.e. $E(A) > E(B)$). This is not to say that both functions are equal, but merely that both are monotonic.

Eccentricity values by themselves cannot define a saliency measure since sites with very little voting strength can produce high eccentricity values. In fact, consider
We are looking for a function that accepts positive vectors as input and results in a measure of the agreement in their orientation. The result should satisfy several criteria:

- The result should be normalized, so that we can compare different sites on a standard scale.

- The measure needs to monotonically increase with the addition of positive contributions.

- It should give higher values to ‘better’ (more directed) spatial arrangements of vectors.

- The effect of proximity should be independent of the effect of agreement.

We can show how the model behaves when a single vector is added to it. Assume the variance-covariance matrix is as follows at state $t$:

$$C^t = \begin{bmatrix} m_{20}^t & m_{11}^t \\ m_{11}^t & m_{02}^t \end{bmatrix}$$  \hspace{1cm} (3.5)$$

The sum of the eigenvalues is the trace of the matrix:

$$\lambda_{\min}^t + \lambda_{\max}^t = m_{20} + m_{02}$$  \hspace{1cm} (3.6)$$

Now adding a new vector $V = [R\cos \theta, R\sin \theta]^T$ to the system will result in a new state $t+1$: 

Chapter 3-Our Approach (2-D) 56
tor weights, and compute moments of the resulting system. Such a physical model behaves in the desired way, giving both the preferred direction and some measure of the agreement. We use the direction of the principal axis (EVmin) of that physical model as the chosen orientation (See equation (3.4)).

\[
\begin{bmatrix}
m_{20} & m_{11} \\
m_{11} & m_{02}
\end{bmatrix} = \begin{bmatrix}
EV_{min} \\
EV_{max}
\end{bmatrix} \begin{bmatrix}
\lambda_{min} & 0 \\
0 & \lambda_{max}
\end{bmatrix} \begin{bmatrix}
EV_{min}^T \\
EV_{max}^T
\end{bmatrix}
\]  

(3.4)

This acts as an approximation to the desired majority vote, without the need to consider the individual votes, but rather the statistics of the set.

![Principal Axis Image](image)

*Figure 3.18 The principal axis of the votes collected at a site is taken as an approximation of the preferred direction.*

### 3.7.3 Saliency measure

The saliency map strength values are taken as the values of the corresponding \(\lambda_{max}\) at each site. So, large values would indicate that a curve is likely to pass through this point. This map can be further enhanced (as shown later) by considering the eccentricity, or \(1 - (\lambda_{min}/\lambda_{max})\). When that value is multiplied by the previous saliency map we achieve better selectivity, and only curves are highlighted. This results in a map defined by \(\lambda_{max} - \lambda_{min}\).
This presents no problem, since all our fields are symmetric around the x-axis, and as such are invariant to the specific rotation. Thus, for \( R \) we get,

\[
R_{a,b} \cdot \vec{V}(i,j) = M_{a,b} \cdot \vec{V}(M_{a,b} \cdot \{i,j\}^T)
\]

Note that for the rotation operator, first the coordinate system has to be transformed, and then the actual vector at that point in space needs to be rotated.

The output of the directional convolution, \( O(i,j) \), is in the form of a variance-covariance matrix for each point in space (see the next section for justification),

\[
O(i,j) = m_{i,j}^{uv} = \begin{bmatrix}
m_{20}^{i,j} & m_{11}^{i,j} \\
m_{11}^{i,j} & m_{02}^{i,j}
\end{bmatrix}
\]

We are now ready to define the directional convolution operator \( EF \oplus I = O \) of a vector field \( EF \) with an input field \( I \),

\[
m_{uv}^{x,y} = \sum_i \sum_j \|I_i,j\|^2 \cdot \left[ \left( R_{I_i,j} \cdot T_{i,j} \cdot EF \right)^{x,y}_x \right]^u \cdot \left[ \left( R_{I_i,j} \cdot T_{i,j} \cdot EF \right)^{x,y}_y \right]^v
\]

where \( 0 \leq u, v \leq 1 \) and \( u + v = 1 \). This defines all elements in \( O(x,y) \).

When the input is a scalar field, the rotation operator becomes the identity operator and has no effect.

Ideally, we would want an averaged majority vote regarding the preferred orientation of a given position. In practice, we treat the contributions to a site as being vec-
3.7.2 Vote representation

The strength of the field is scaled by the contrast of the segment, so that stronger segments have stronger votes throughout the space.

Note that, although the process is local in essence, the fields impose some global order, and one line segment can implicitly 'vote' for a large curve without any explicit global reasoning involved.

As before, we define the Extension Field, \( EF(i,j) \) at any site \((i,j)\), to be a vector field:

\[
\overrightarrow{EF}(i,j) = (EF_x, EF_y)^T
\]

The input \( I \), will be denoted with a bar when oriented (a vector field), and without when non-oriented (a scalar field).

We define the translation \((T)\) and rotation \((R)\) operators to translate and rotate an arbitrary vector field \((V)\) respectively. For \( T \) we have,

\[
T_{\Delta x, \Delta y} \cdot \overrightarrow{V}(i,j) = \overrightarrow{V}(i + \Delta x, j + \Delta y)
\]

which simply translates the indices of the target vector field.

For \( R \), we first define a rotation matrix \((M_{a,b})\), that brings the unit vector \((0,1)^T\) to any given unit vector \((a,b)^T\). Such a matrix always exists, but may not be unique.
ures 3.16 and 3.17 is the way vectors from the fields interact. In 3.16, sites along the straight line have perfect agreement in orientation, while sites off the main line show high degree of disagreement in orientation.

The next step is to represent these accumulated vector votes at each pixel in a convenient and meaningful way. We argue that the relevant information is computed simply by observing the distribution of these votes, and more specifically well approximated by the best fit ellipse representing the moments of these votes. We give the mathematical details below.

Figure 3.17 Superposition of two Extension Fields over a scenario of two co-circular short segments (dark lines).
3.7 Computation of the Saliency Map

3.7.1 Vote Accumulation

Having designed the field for a given input site (from a dot to an edgel) we superpose these votes from each active site at each location. A pixel receives a vote from all sites whose field overlaps this pixel, and each vote has the form of a vector, given by its strength (length) and orientation.

For each voting site, we align the appropriate extension field with the direction of the site, center the field at the site, and compute the vector contribution at each location covered by the field. This is repeated for all input sites. This process is similar to a convolution with a mask, except that the output is a vector instead of scalar. The procedure is illustrated in Figures 3.16 and 3.17. The important thing to notice in Fig-

![Figure 3.16 Superposition of two Extension Fields over a scenario of two co-linear short segments (dark lines).]
some pre-defined angle. The actual angle is of course a function of the amount of error one wishes to allow for a straight line. (Figure 3.14).

A second, and better motivated method would be to generate a new field by convolving with a short straight line. This last method is more in the spirit of our approach, and also retains the notions of feature fuzziness and continuum of saliency values. Again, the length of the straight line with which we convolve, will determine the amount of desired error in our definition of ‘straightness’. A typical Straight field is depicted in Figure 3.15.

Figure 3.14  One way of constructing the straight line Extension Field.

Figure 3.15  A Straight Field generated by convolving the Extension field with a straight line.
By adding additional constraints it is possible to construct a saliency operator to enhance all straight line formations in the image, and, at the same time, suppress other smooth curves.

We now derive the shape of a field capable of finding all straight, or almost straight line formations in a cluttered directional edge image. This can be done in two ways.

The first method would just use the part of our original Extension Field that applies to straight lines. This simply means cutting off all field elements that are outside Figure 3.13 The shape of a semi-deterministic field. Such a field encodes some degree of uncertainty in the orientation of the input edge (+/- 1 radian).
would like to treat the image as if it was made out of non-directional tokens (or dots), and apply the point field to it.

### 3.6.3 Unifying the field concept

We unify the maximum certainty (Extension) field and the Point field (and all fields in between) by considering a continuum of eccentricities associated with the multi-directional edge. The Extension field is constructed from a degenerated instance of an ellipse (a line segment), while the Point field is created from a ‘circular’ edge point. This makes full use of our input model and allows treatment of mixed images in a consistent and unified way (see Figure 3.12). As an example we show how a field with a +/- 1 radian of uncertainty would look like (see Figure 3.13).

In practice, it was found that the Extension field, although designed for perfect orientations, tolerates some degree of uncertainty in the input samples. This is due to the special interpretation of the votes, which is described later.

### 3.6.4 Special Purpose Fields: The Straight Field

Special purpose fields are fields synthesized to enhance a special feature in an image. For example, in aerial images, a desired feature could be rectangular rooftops.

---

8. i.e. having varying certainty measures (or a mixture of dots and lines)
A suitable field must have a circular symmetry, and in practice is constructed by convolving our original extension field with a ‘multi-directional’ edge segment, as shown in Figure 3.11(b-c).

The Point field’s (PF) equation now becomes

$$PF(x) = e^{-Ax^2}$$  \hspace{1cm} (3.3)

where $x$ is the euclidean distance from the origin.

A typical input is shown in Figure 3.11(a), where a broken sine wave and a set of random points on a circle are embedded in noise.

Another scenario where the point field could be useful is in images where co-curvilinear formations between features other than dots are present. In such cases we

7. convolution of fields is defined in the next section.
3.6.2 Other fields

The Point Field

A dot image is a degenerate case of an edge image, where the edges have no direction. Such maximum orientation uncertainty in the input fits our input model well, and allows us to handle such cases in a uniform way. Perception is weakened by the loss of orientation data, and we are only able to handle cases with a moderate amount of noise. The only applicable perceptual law is that of *proximity*. Figure 3.10 demonstrates how removing orientation data (by converting the edgels to points) weakens perception of the salient circle.

*Figure 3.10*  Loss of orientation $\Rightarrow$ Loss of perception. The segments of (a) were replaced by dots in (b). The perception of the circle is weakened.
where \( x \) is the distance along the circular arc and \( \rho \) is the curvature of the given arc.

The parameter \( A \) controls the proximity decay, while \( B \) controls the decay due to higher curvature. The parameters \( A \) and \( B \) were selected empirically based on the above-mentioned constraint. They are not however, independent, and changing one will require a change in the other, to preserve the condition of maximum undecidability. The shape is thus that of a decaying exponential, where the decay is a function of both the distance and the radius of the corresponding circular arc. The parameter \( A \) is related to the size of gaps we are able to recover, and should be derived from image size and sparseness of data. In our implementation, \( A \) was selected to bridge gaps of up to about 50 pixels (\( A=0.003 \)), and \( B \) was adjusted to loosely satisfy the above-mentioned constraint (\( B=2.85 \)).

In a recent work by Kardaras and Medioni [35], an evaluation of the influence of these parameters on the quality of the saliency maps, was performed. It appears that by selecting faster decaying parameters, better selectivity is achieved. Selectivity is the ability to discern two close-by lines. However, a fast-decaying field is not able to bridge large gaps, since the contribution far away from the center of the field tends to be of the same order of magnitude as the numerical precision of our implementation.
This scenario, we claim, is a middle point between a clear choice of a connection by a sharp junction, and a connection by a smooth curve. We regard this to be the most competitive scenario in terms of grouping of the two line segments. We thus assign probabilities to the field elements in such a way that all paths connecting these segments are assigned roughly the same saliency, and there does not exist any single best path between the two. More precisely, we set the field element strengths such that all values within the marked triangular region are the same. Such a scenario, when repeated for all distances, removes all but one degree of freedom as to the choice of values for the field. The weights will not in general correspond to any analytical function. In fact, it is not possible to precisely satisfy this constraint along the straight lines emanating from the two edgels.

For practical reasons, therefore, we have decided to ‘guess’ an approximating analytical function to the Extension Field (EF) and optimize its coefficients. We set the decay to be of Gaussian nature for both the proximity and the curvature constraints, as shown in equation 3.2.

Figure 3.9  5 arrangements of two segments, each with a different separation angle. Angles much smaller than 90 degrees suggest a corner, while angles much larger than 90 degrees suggest a smooth connection.
Another reason all values beyond the two main diagonals are zeros, is a technical one. Having a segment vote for a point in space which is more than 90 degrees away (along a circular arc) could potentially cause unrelated segments to vote for the same curve, even though such a curve should not connect them.

### 3.6.1.2 Strength

The main consideration here is that we would like a decay due to distance *and* a decay due the higher curvature. Many functions can fulfill this, and the problem is under-constrained. The best way to determine these values is by considering an *intentionally* ambiguous or *undecidable case*. The assignment of actual probabilities to the field is thus performed as follows: We consider two short edge segments, *perpendicular* to each other and equally distant from the origin\(^6\) (see center of Figure 3.9).

---

\(^6\) This scenario is termed ‘*the maximum undecidability arrangement*’.
This is the integral of the curvature (brought to some power) along the curve, where $\theta$ is the tangent along the curve $\gamma$, and $s$ is a parameter ‘running’ along the curve. The variable $\alpha$ is traditionally taken to be equal to 2, but it can be shown (See “Mathematical Treatment (Proofs)” on page 165) that the choice of a circle as the connecting curve in the scenario shown in Figure 3.7 minimizes the TC for all values of $\alpha$ greater than 1 [24]. A larger $\alpha$ would just penalize sharp turns more than a smaller one. We thus set the orientations of the field elements to be tangent to a circular arc connecting the origin and that point in space. Note the for a straight line the total curvature is always 0.

It is important to note that an edge and a point cannot always be joined optimally by a circular arc. Such pairs may be subject to some more complicated smooth curve connection, but we have decided not to handle explicitly these cases. It is quite obvious (see Figure 3.8) that extending a curve beyond the 90 degree point does not satisfy the minimum curvature constraint.\(^5\) For this reason our Extension field has zero values above and below the main diagonals (as is evident from Figure 3.6(a)). This merely means that we choose not to vote for pixels in that area, and additional ‘in-between’ information is necessary to reconstruct curves between such pairs.

\(^5\) Note that our proof holds only in the 0 to 90\(^0\) interval.
3.6.1.1 Shape and Orientation

Since we favor small and constant curvature, field direction at a given point in space is chosen to be tangent to the osculating circle passing through the edge segment and that point, while its strength is proportional to the radius of that circle. Also, the strength decays with the distance from the origin (the edge segment). The choice of a circular extension agrees with the constraint of smallest total curvature. For points in space that are along the x axis, the osculating circle degenerates into a straight line (or a circle of radius infinity).

In trying to computationally evaluate the various constraints over a given curve we find that a (somewhat revised\(^4\)) measure of Total Curvature (as used by [61]) encompasses most of the desired constraints. We define the Total Curvature (TC) to be:

\(^4\) In [61], \(\alpha=2\) is used, and the absolute value becomes redundant.
In other words, it votes on the preferred direction and the likelihood of existence of every point in space to share a curve with the original segment. The field is of infinite extent, although, in practice, it disappears at a predefined distance from the edge. Figure 3.6 depicts the Extension Field.

![Extension Field](image)

*Figure 3.6 The basic Extension Field. (a) Direction, and (b) Strength.*

The design of this field needs to account for the shape of the field, the orientation at each site, and the strength at each site. The first two aspects are quite straightforward, while the last needs more explanation.

### 3.6.1 Design of the Extension Field (Orientation and Strength)

The design of the field involves the determination of orientation of each field element and its strength.
Consider however figure 3.5. Grouping based solely on the tangents of end-

![Figure 3.5](image.png)

*Figure 3.5 Both scenarios have same first order compatibility figure. We would, however, want to group the left curves, but not the right.*

points would either group or not group both instances, subject to some arbitrary threshold. We want our scheme to create a stronger group of the left pair but not the right one, based on a more global co-curvilinearity criteria.

It is not reasonable to expect that extension curves from two different extension fields will align throughout their extent. It is more likely that such extensions align locally in many places. For that reason, the extension field will comprise of local best candidates for extensions. In the next section we define the exact shape and usage of the Extension Field.

### 3.6 The Extension Field: design and implementation

**Definition:** An *Extension Field* is a non-normalized probability directional vector field describing the contribution of a single unit-length edge element to its neighborhood in terms of length and direction.
of some kind of a field (or flow) emanating from the end-point of a curve. The idea of an extension field plays major role in our scheme, as we will show later.

### 3.5.3 Best Connection between Two Line Segments

The situation occurs whenever a ‘compatibility figure’ of two close-by segments is computed in order to decide whether these segments should be grouped together. When determining the compatibility of two lines we would like to consider for each line its best extension (or extension field, as discussed in 3.5.2) and arrive at some compromise as to the best path between them. In other words, each of the curves votes (using its field) for a family of curves. If the curves should really be connected then some extension from curve 1 would align with another extension of curve 2, to form the compromise. The idea of a compromise between extension fields is also central to the approach, and will be presented in a more formal way in later chapters.

---

2. A measure of ‘agreement’ between two curves, based on the difference of the end-point tangents and separation. (as used by [17, 42, 47])
3. Which is not necessarily the best continuation of any of the existing curves, or even a connection between the given end-points.
Many approaches in the past (e.g. [17, 42]) used the tangent of the end-point to determine the best extension. This approach cannot always work properly for three reasons:

1) The tangent is very sensitive to noise and may introduce large errors.
2) The end-point may not be determined uniquely if the curve in question is fragmented.
3) The extension can only be a straight line (first order), thus not taking into consideration the global shape of the curve.

We would thus want to consider all of the curve in such a way that the extension is smooth, influenced most by the behavior of close-by points of the curve, but can assume any form. Note that we would also like to be able to construct the extension even if the curve is fragmented.

We go further than that. Rather then having only the best extension, we would like to list all possible extensions in the order of their likelihood. This suggests the use of a method that can produce a variety of extensions, each with a different level of confidence.
3) Favoring low curvatures over large ones - Humans seem to connect fragmented line segments in a way that the increase in total curvature is minimized (see Shashua Ullman [61]).

4) Proximity - Closer segments influence each other more than distant ones.

The claim is that these properties are invariant to point of view, and are unlikely to appear by accident. Co-linear lines in the image are likely to come from true co-linears in the scene, and the same is true for co-curvilinear lines.

We have not mentioned grouping laws due to symmetry and to end-point formations. These will be discussed in later chapters.

With that in mind, we have devised a technique that implicitly imposes the above constrains in the form of an Extension Field emanating from each edge segment, as discussed in the next few sub-sections.

### 3.5.2 Extending a Curve

Given a curved line (as in Figure 3.3) we ask the question: What is the shape of the most ‘natural’ extension, based on the mentioned constraints?
should assign a very low value to areas of no saliency. Furthermore, curves with strong
saliency should be assigned larger values then weaker curves.

Note that saliency has not been defined in any precise way so far, and when de-
scribing the model of the output we aim at the ideal case, where the saliency map cor-
responds to our subjective judgement of what is salient and what is not.

3.5 Rationale for the Extension Field

In order to define saliency qualitatively, we start by writing down the major con-
straints which govern our mechanisms of saliency (at least according to the Gestalt
theory).

3.5.1 The perceptual constraints

The underlying goal is to keep the interpretation as simple as possible in the 'Ge-
stalt' sense. This translates into four major constrains:

1) Co-curvilinearity - In the lack of other cues, smooth continuation is the only
   interpretation, and so is co-curvilinearity. (This is referred to as ‘good continuation’
   in psychology literature).

2) Constancy of curvature - We tend to extend a curve of some constant curvature with
   the same curvature, keeping the interpretation as simple and regular as possible, yet
   consistent with our sensory information. This principle is called \textit{Prägnanz} by
   Gestalt psychologists (see Figure 3.2).
Uncertainty with respect to orientation is inherent in the process of edge detection since, in many cases, a discrete set of oriented masks are convolved with the image, and the one with the largest response determines the direction of the local edge. This value is accurate only up to the resolution of the set of masks, and can be used as an uncertainty measure in our scheme.\textsuperscript{1}

Directional certainty is directly proportional to our ability to extract useful features, because the information content is reduced when uncertainty is introduced in the input. Consider, for example, Figure 3.10 on page 46. Figure 3.10(a) has maximum directional certainty, while Figure 3.10(b) has none. Clearly, it is much harder to extract useful features from the latter.

### 3.4 Model of the Output

Our model of the output is related to Sha’ashua and Ullman’s [61] in the sense that a \textit{saliency map} is first constructed from an edge image, and higher-level features are inferred later. Our saliency map assigns a value and a direction to \textit{every} site in the image.

Ideally, such a saliency map should assign large values of likelihood along illusory lines (as well as along physical curves), and also specify a direction of most probable continuation of any given segment. Such a map will enable us, at a later stage, to group features by following the salient connections between the primitives. The map

\textsuperscript{1} Recently, steerable-scalable masks (e.g. [53]) claim a much more accurate detection of orientation (1 degree resolution) away from corners.
3.3 Model of the Input

We would like to associate with each site of an image a direction, strength and a degree of uncertainty for that direction. This can be nicely approximated by an ellipse (as illustrated in figure 3.1). By using such an input model, one site could be classified as being a part of a curve with known orientation and no uncertainty, while another as being a point with completely uncertain orientation (the ellipse becomes a circle). Thus a circular site could be interpreted as a junction. No explicit provision is made here to describe intersections between curves, or the local incident angles of such junctions. It is also the case that such information is rarely available at the input.

We can thus use as input either the thresholded output of any edge detector (with no linking) or even an un-thresholded version of the edge detector output. It will be shown that our system yields almost the same results with different choices of this threshold, as long as a sufficient number of useful features are present (see section 3.9.3.5).

Figure 3.1 A simple input site model. Every site is associated with a preferred direction, strength and eccentricity (or uncertainty).
Our system can take as input a set of oriented features, as produced by a typical edge-detector, as well as features with an associated directional uncertainty, and even edges with no associated orientation (dot formations). The output of our system is a set of oriented features, with an associated strength reflecting its saliency, and an uncertainty of orientation.

3.2 Philosophy

A serious deficiency of most grouping schemes is the fact that arbitrary thresholds are used. These thresholds rarely have a theoretical justification, and in many cases are based on experimentation. These thresholds restrict the scope of operation of the operators used, and as such commit to certain grouping decisions which may prove to be wrong later on.

In our scheme we delay decisions as much as possible. For this reason, all measures are continuous, and in a sense wrong groupings are correct to some degree (though probably a small degree). We make extensive use of external properties of the data, which do not require a threshold, and enable us to localize or thin results. Furthermore, the output of our algorithm is a sorted list of groupings from the most likely one to the least likely. It is up to other processes to decide how much of this list they wish (or can) use.
In our method, every site (pixel or other size cell) collects votes from each segment in the image. These votes contain orientation and strength information preferred by the voting segment. A measure of ‘agreement among votes’ (in terms of orientation) is computed, and sites which have high agreement values are considered salient.

In more technical terms, a vector field, referred to as the Extension Field, is generated by each segment, and a function over the whole space determines points of saliency. A subsequent step links areas of high saliency to produce a description in terms of curves and junctions.

Our voting scheme is somewhat related to the Hough transform approach [30], but can detect shapes defined by their properties (smoothness etc.) rather than by their exact analytical shape (lines, circles, etc.). A study of the relations between our scheme and the Hough Transform is given later in this chapter. (See “Comparison With the Hough Transform” on page 86.)

The process is likely to produce features more similar to what we expect, both in terms of saliency and connectivity. It completely eliminates the critical influence of threshold values, and since noise is not likely to produce high ‘agreement’ values, it gets attenuated.

We believe that the use of a non-linear global voting scheme is fundamentally different from the many local techniques reported before. Such a non-linear global approach cannot directly be decomposed into an iterative local scheme.
Chapter 3

Our Approach (2-D)

3.1 Overview

As was presented before, the physical evidence extracted \textit{locally} from images (e.g. through edge detectors) is in many cases ambiguous and does not correspond to the expected perception of the image. It is thus necessary, we believe, to impose \textit{global} perceptual considerations even during (or as part of) the low-level processing. These constraints need to be generic, as they should embody expected properties of \textit{all} scenes.

We briefly explain below what is computed, how it is computed, and give a rationale for the approach.

In signal edge-detection, we extract edges and assign position, orientation, and strength based on local physical measurements. Unfortunately, the relationship between the strength of the perceived edges and the strength of the measured signal is not straightforward. We therefore propose to generate a different measure of strength, which we call saliency, based on more global expected properties of the observed scene, such as co-curvilinearity.