Families of models for gabor paths demonstrate the importance of spatial adjacency

Roger Watt
Department of Psychology, University of Stirling, Scotland, UK

Tim Ledgeway
School of Psychology, University of Nottingham, UK

Steven C. Dakin
Institute of Ophthalmology, University College London, UK

This paper reports psychophysical and modelling results concerning the contour-detection paradigm of D. J. Field, A. Hayes, and R. F. Hess (1993). We measured psychophysically the maximum tolerable contour curvature (path angle) as a function of contour length. We compared these data to the predictions of an association field (D. J. Field et al., 1993) model based on the relative positions and mutual orientations of nearby elements and to models that explicitly link adjacent elements into chains and characterize each chain by its sequence of contour bends. For every stimulus, a large set of chains is produced and the target identified as the chain with the lowest maximum bend. We tested two different types of linking process: isotropic (linking one element to any other nearby) and anisotropic (linking one element to any others nearby along the orientation of its axis). All of these models can account for our data. Moreover, we show that the pattern of results due to path angle is principally a product of the distribution of spurious contours in the randomly oriented background. Given that some of the models do not embody constraints of orientation relationships between linked elements, this finding shows the importance for early vision in deciding which local elements are to be associated.

Keywords: spatial vision, computational modeling, shape and contour


Introduction

The initial stages of visual analysis are mediated by neurones that are sensitive primarily to structure falling within a restricted region of visual space known as their classical receptive field (e.g., Hubel & Wiesel, 1968, 1977). For example, in primate visual cortex, neurones in area V1 representing central (foveal) vision have orientation-selective classical receptive fields that often cover less than 1.0 deg². Optical imaging studies (e.g., Bonhoeffer & Grinvald, 1991) have revealed that primary visual cortex generates a retinotopic (local) representation of contour orientation. The local image structure that best drives V1 neurones is typically simple in spatial form: perhaps even as simple as to be spatially well approximated by local elongated structures varying in just orientation and spatial scale. If this is the case, then a set of templates (e.g., filters or receptive fields) can be used to create image descriptions of such local structure.

These strictly local measurements must be subsequently linked, as appropriate, across space to encode the overall structure of spatially extensive objects present in the visual scene. Since these larger scale overall structures are essentially unconstrained, the linking processes must be similarly unconstrained. This is important, that the process is capable of generating completely novel outputs (Ullman, 1984; Watt & Phillips, 2000), as it indicates the need for processes that are not readily modelled as templates.

Longer range interactions between local sensors have been implicated in a wide range of perceptual grouping phenomena. These include global stereoscopic form detection (Julesz, 1971), global motion perception (Chang & Julesz, 1984, 1985; Williams, Phillips, & Sekuler, 1986), contextual modulation of contrast sensitivity (Polat & Sagi, 1993), image segmentation (Kovács & Julesz, 1993), and detection of extended motion trajectories in noise (Bex, Simmers, & Dakin, 2003; Verghese, Watamaniuk, McKee, & Grzywacz, 1999).

The issue of how local orientation measurements are combined across space to describe extended spatial contours has received considerable interest in the last decade or so, driven in no small part by the development of the contour-detection paradigm of Field, Hayes, and Hess (1993). Here, the observer’s psychophysical task is to detect the presence of a smoothly curved contour (path), composed of a set of spatially separated oriented Gabor patches (elements), in an array of similar but randomly oriented background elements. Performance on this task depends on many stimulus parameters that have been summarized elsewhere (for reviews, see Hess & Field, 1999; Kovács, 1996), but the main findings are described below.
The most widely considered stimulus parameter is the contour curvature (path angle): the change in orientation between adjacent Gabor elements. For example Field et al. (1993) report that detection performance is best (~100% correct) for straight paths and declines gradually as the degree of curvature of the target contour increases, reaching chance levels for path angles greater than about 60°. However, this upper limit on performance can vary between 30 to 60° depending on various stimulus factors.

Performance on the psychophysical task is highly dependent on the density of elements, at least for moderately and highly curved contours (Li & Gilbert, 2002; Pennefather, Chandra, Kovacs, Polat, & Norcia, 1999), and fails at inter-element separations of about 4 to 6 times the Gabor wavelength (Kovács & Julesz, 1993).

Observers’ ability to detect the presence of a target also depends upon the orientation of the Gabor elements with respect to the local orientation of the contour they form (Bex, Simmers, & Dakin, 2001; Field et al., 1993; Ledgeway, Hess, & Geisler, 2005). Performance is best when the element orientations are aligned with the local contour orientation (snakes), moderately worse when they are orthogonal (90°, ladders), and hardest to detect when they are oriented obliquely (45°) with respect to the contour. All three stimuli provide the same basic statistical quality of cue for observers—that snakes are nonetheless easier to detect than the others suggests that the mechanism involved is specialized for contour integration.

Field et al. (1993) interpreted their original results in terms of an “association field” that combines responses from neighboring local filters or receptive fields tuned to similar orientations. The notion is that there is an association field about each element that produces a scalar association output that increases with the amount of nearby correctly oriented and aligned contour structure. The strength of the association is greatest when there are nearby elements that are collinear and reduces with increasing distance, curvature, and misalignment from co-circularity. Stated in this way, the association field does not make explicit the relationship between elements and does not identify contours. However, elements that are part of a smooth contour will tend to have higher associations.

There are, however, a number of potential limitations to the association field concept as a general model of contour description in vision. First, the structure of the association field may be over-constrained. It is essentially a template for stimuli that have been found to be detectable and as such may be more descriptive of the results of experiments rather than explanatory in nature. There could be many other templates (not necessarily so closely modelled on psychophysically detectable stimuli) that are equally able to distinguish smooth contours from noisy backgrounds by virtue of some combination of the differences between contour and background.

Second, one fundamental property of the association field, as originally conceived by Field et al. (1993), is that adjacent elements are strongly linked only if they satisfy the joint constraints of position and orientation along smooth curves. There is evidence that contours are readily detected when defined by other types of position and orientation relationship. For example, a single association field cannot explain the (non-monotonic) patterns of performance found when the orientations of the Gabor elements along a path are systematically misaligned with respect to the axis of the contour they depict (e.g., Ledgeway et al., 2005). The notion of composite association fields, simultaneously sensing multiple different contour forms, may appear to deal with this but such models cannot subsequently distinguish which pattern they have responded to. The consequence is that multiple association fields with different orientation and positional properties (e.g., Yen & Finkel, 1998) become required—essentially further templates.

Although there is some suggestion that the shape of contours in natural images may be statistically constrained somewhat (Geisler, Perry, Super, & Gallogly, 2001; see also Sigman, Cecchi, Gilbert, & Magnasco, 2001), there is in principle no limit on the form of a contour in an image. If the range of possible contour shapes is unlimited, then template methods of contour description are difficult and generative descriptive methods are more appropriate. A generative process is like a language: it uses a small number of symbols (say representing local piecewise straight segments of contours) and combined by a small number of methods (such as connected to) but in unlimited numbers to have an infinite descriptive capacity. Any continuous line can be reasonably described by a concatenation of local pieces. Watt (1991) has shown a form of image description language that is of this type.

Consequently there is a need to consider alternative schemes that may also be able to account for human performance measured with the path-detection paradigm. The local output of an association field is a measure of how good the relationship is between the element at that point and any others that fall within the association field range. In effect, it combines two logically distinct operations: linking an element to those around it and simultaneously assessing the goodness of the resultant contour configuration of all linked elements. In what follows, we compare this type of model with a link-then-describe type of model.

In a link-then-describe model, the two types of spatial relation, relative position and relative orientation, could be used to constrain either the link or the description. The aim of the present study is to investigate the separate contributions of each of these two processes to performance. The results of psychophysical experiments are compared to the performance of three different families of model, each using constraints based on a different combination of element position and orientation information. The
comparison illuminates the relative significance of the different information sources. Specifically, the models considered are

i. an association field model which responds most strongly to elements surrounded by further contour with appropriate combinations of position and orientation;

ii. models that link each element with others solely on the basis of spatial separation and then select linked groups of elements with the lowest curvature; and

iii. models that link an element with others on the basis of spatial separation and the spatial direction of those others.

The first family of models, association fields, have selectivity for both where elements are with respect to each other, and what form their combination makes: a part of circle or something similar. The other two families only have selectivity for where elements are with respect to each other, regardless of the resultant form. In the light of the computational need for generativity and to be able to recognize a wide range of forms, the second and third families of models might be computationally more desirable.

To anticipate the results we find that all models exhibit similar performance to human observers, despite their qualitative differences. In this regard there is little empirical basis to select one type of model over another, at this stage, other than a general preference for the more simple models.

### Psychophysical data

#### Observers

We used 4 observers, all with normal vision. One was an author (RW), while the other 3 were naïve to the purposes of the experiment.

#### Apparatus

Experiments were run under the Matlab programming environment (MathWorks Ltd) on a PC computer. Stimuli were displayed on a Sony CRT monitor operating at 75 Hz, whose luminance had been linearized in software. The mean luminance of the display was set to 25 cd m$^{-2}$.

#### Stimuli

Images were presented within a square display window (568 pixels in each dimension) that subtended 11.3° at the viewing distance of 1 m. Each stimulus-element was a Gabor patch in cosine phase, with wavelength $\lambda$ (0.12°), Gaussian envelope of $SD \lambda/2$ (0.06°), and Michelson contrast of 50%.

A path was defined by a sequence of element locations and orientations. The distance from one location to the next varied according to the same distribution as the separation between neighboring elements in the background composed of randomly placed Gabors. In a group of 3 elements, the change in direction from the line joining elements 1 and 2 to the line joining elements 2 and 3 is called the path angle. For any one target the magnitude of the path angle was fixed, although its sign changed randomly after a minimum of three elements. The orientation of the element at each location was set to the orientation of the chord joining the elements on either side. This resulted in the paths having a small degree of departure from co-circularity.

The path was placed in either the left or the right half of the stimulus field. To complete the stimulus, randomly oriented elements were repeatedly added at randomly selected locations—prescribed to lie so that the center of the new element was not less than 2.5$\Delta$ (0.3°) from the center of any other element—until no more elements could be added. This results in a mean distance between adjacent elements of 3.3$\Delta$. We will use this mean inter-element spacing as the basic unit of distance, 1Δ. The whole field was of size 38 Δ and could accommodate about 1000 elements. Since background elements were added at random locations, the distribution of distances from one element center to those of its immediate neighbors is also random. The minimum distance is 0.75Δ; the mean distance is 1Δ and the maximum is 1.5Δ.

#### Procedure

The experimental protocol was designed to limit knowledge of the stimulus available to the observer, either by instruction or by observation of the stimulus set. This more closely matches the normal demands of vision where the lengths and forms of contours are typically unknown to the observer. To this end, a run of trials comprised stimuli of varying path angle and varying path length, in a random sequence: on any one trial the observer could not predict what the specific stimulus parameters were.

In all cases, the instructions to the observer were to detect the most salient smooth contour and to indicate, using the computer keyboard, whether it fell in the left or the right half of the stimulus. Stimuli were displayed until the observer made a response. The duration of the stimuli was unconstrained for the following reasons. First, we sought to measure performance under conditions that would be both optimal and naturalistic for the observer and to enable a fair comparison to be made with the performance of the models. Each of our models operates on a purely spatial representation of the image and thus
presentation time is irrelevant in this context. Second, curtailing the duration of the stimulus is known to degrade performance on this task, due to extraneous factors that may have little to do with the ability to link elements in the stimulus. For example, Field et al. (1993) found that detection deteriorated when the duration of a path stimulus was reduced from 1 s to 0.25 s, but the drop in performance was relatively uniform across the range of path angles tested. They point out that when fixation is unconstrained, limiting the stimulus presentation time can severely restrict the number of fixations that can be made to foveate the path when its location is unpredictable from trial to trial (as it was in the present study).

Path angle varied in 9 steps of 5° from 0° to 40° (inclusive). Twenty examples were generated for each required stimulus type, and then used once in each run. Since a typical run of trials would involve as a minimum 6 path lengths × 9 path angles × 20 examples (1080 stimuli), observers did not notice the repetition of stimuli on the next run of trials.

For the basic analysis, the data from each condition are presented as a function showing percent correct detection performance with increasing path angle. Correct responses typically decline from 100% to 50% (chance) as path angle increases, and a smooth curve was fitted to the data in order to characterize overall performance. The curve was a simple cumulative Gaussian, with a third parameter, an exponent on the stimulus level (path angle). The path angle producing 75% correct responding (referred to as the maximum tolerable path angle) was obtained from the fit and 95% confidence limits on this estimate were calculated using a bootstrap method.

Results

We first measured contour detection as a function of the number of elements in the path target. As can be seen in Figure 1, longer path lengths lead to larger maximum tolerable path angles. Paths of length 3 are not readily detectable at all.

The maximum tolerable path angle for each path length is shown in Figure 2. Performance is generally lower than reported elsewhere (e.g., Field et al., 1993). However, their stimulus differed in many respects, especially in that they used fewer background elements and path lengths of 12 and constrained paths to fall within a maximum distance from the center of the display. To assess these differences, the first measurements were repeated with two field sizes, 1000 elements (as before) and 250 (which is broadly similar to Field et al., 1993). In addition a wider range of path lengths (up to 16) and path angles (up to 60°) were investigated. The results of this second set of measurements are also shown in Figure 2. Perhaps unsurprisingly, the smaller field leads to higher performance: a fourfold reduction in the number of background elements.
elements likely leads to the presence of fewer spurious targets.

Previous work in this area has indicated that performance reaches asymptote for 7 elements for dot stimuli (Moulden, 1994) and 10 elements for Gabor stimuli at 200 ms (Braun, 1999). Undoubtedly these variations are due to features of the experiment such as field size.

**Simulations**

Stimuli were created exactly as if for a psychophysical experiment. The length of the target path varied from 3 to 8 elements, and the path angle varied exactly as in the first experiment. For each combination of stimulus parameters a response was calculated for the model, indicating on which side, left or right, the most likely target lay. The model response was obtained blind to the stimulus parameters of path angle and path length. Exactly as in the psychophysics, percent correct as a function of path angle is then obtained. In all cases, this has the same general form as the human data: a decline in performance with increasing path angle.

**Model family 1: Association fields**

In this family of models, each element in the image is surrounded by an association field which is used to determine how likely it is that the element belongs to a continuous contour with other elements. The likelihood is higher when the association field is stimulated by other elements at positions and orientations that would be close to co-circular.

**Form of the association field**

We have implemented a simple form of association field using the product of two types of weight structure centered and aligned in orientation with each element: a relative position function and an overlapping relative orientation function. In this way, an association field is selective for both for where elements lie and what form their combination creates.

The relative position (“where”) weighting function, concerned only with the position of the second element, is the product of a distance function, which preferentially weights closer second elements, and a curvature function which preferentially weights second elements whose positions lie on lines from the origin element of lower curvature. The distance function is a Gaussian parameterized by a linear extent $\sigma_d$. The curvature function is also a Gaussian parameterized by a rate of orientation change $\sigma_c$.

The relative orientation (“what”) weighting function applies highest weight to co-circular orientations and declining weight to orientations that deviate from co-circularity. This function is a Gaussian in orientation, centered at each position on the co-circular orientation at that position and parameterized by an orientation error extent $\sigma_t$.

Figure 3 shows an association field, as a function of space and orientation. It is centered on an element at its origin. Any other element lying inside the propeller-shaped surface will generate a high weighting at the origin. A projection of the function on to the 2D plane shows the “where” selectivity. A change in the distance parameter, $\sigma_d$, just changes the overall size of this selectivity. A change in the curvature parameter, $\sigma_c$, changes the extent to which this “where” selectivity extends away from the line $y = 0$. The vertical dimension of the space corresponds to the “what” selectivity of the association field. At any one point in $x$--$y$ space, the
Mechanism of the association field

The overall association strength at any one element in the field is made simply by summing a set of individual association weights, one arising from each of the other elements (although most of these weights will have a negligible value). Thus, those elements that lie in a good relation to their neighbors will have a high association strength. The application of an association field to a whole stimulus produces a set of scalar continuous-valued responses, one for each element. This procedure does not explicitly identify the contour or any quality of the contour.

We have adopted a simple approach to decisions about the outputs of association fields by simply supposing that the most salient contour in the image will include the Gabor element with the greatest total association strength. In line with the psychophysical procedure, a response from the model to a specific stimulus is obtained by finding the side of the stimulus field, left or right, that contains the highest association strength.

Simulation 1: Effects of varying the three parameters

In the first simulation, the target was a path composed of 8 elements with path angles that varied from 0° to 40° in 5° steps (as in the psychophysical experiment). All 3 model parameters, \( \sigma_d \), \( \sigma_c \), and \( \sigma_t \), were systematically varied. For any given set of values for the 3 model parameters, the proportion of targets correctly detected was obtained as a function of path angle. This function is not shown, but it exhibits the same decline in performance with path angle as the psychophysical data. From this function, the path angle supporting 75% correct performance was obtained—the maximum tolerable path angle. The 3 parameters were all independently varied and for each combination of parameters, the maximum tolerable path angle was measured. This results in a 3-dimensional performance space. The top panel of Figure 4 shows an iso-surface in these dimensions that encloses all parameter combinations that are as good as the human data or better.

For further analysis, we present the results in 3 separate ways. First, the optimum performance is found for \( \sigma_d = 1.0 \), \( \sigma_c = 35 \), \( \sigma_t = 7 \). At these values, the maximum tolerable path angle is found to be 45.6°, compared with a mean value of 32° for the observers. We note that the parameter values are broadly in line with the equivalent stimulus values.

Second, to assess how performance falls off away from this optimum, the results are shown as three individual functions in Figure 4 (middle panel). In each case, 2 parameters are set to their optimum while the other one varies. Small values for \( \sigma_d \) lead to very low performance because the association field does not stretch adequately from one element to the next. Larger values show a gradual decline in performance arising because larger distances include more elements, some of which will have spuriously good relationships. Small values for \( \sigma_c \) result in poor performance because this parameter sets a limit on the maximum bend that will be responded to. Interestingly, larger values show little if any decline in performance, suggesting that this parameter is not particularly important. Similarly there is very little consequence of variations in \( \sigma_t \).

Third, to assess how the parameters interact with each other, Figure 4 (bottom panel) shows the variations in performance for each possible pair of parameters, with the third parameter set to its optimum value. A contour is drawn on each which contains all the parameter values where performance is equal to or exceeds the human performance. As can be seen, this space is very broad, indicating that the individual parameter values are not very important in determining the performance of the model and are therefore largely unconstrained by the psychophysical data.

Simulation 2: Effects of path length

The psychophysical results show a strong effect of the length of the path, at least up to 8 elements. Simulations were run with various combinations of model parameters to explore this. In each case, the whole psychophysical
procedure was repeated, with path lengths between 3 and 8 and path angles between 0° and 40° in steps of 5°. In common with the psychophysical procedure, the model did not know what path length or path angle to expect on any given trial. Performance was assessed with 7 different combinations of the 3 association field parameters. The combinations chosen were (i) the optimum value for each of the parameters; (ii–iv) for each parameter, the largest value consistent with performance at the human level and with the other two parameters set accordingly; and (v–vii) for each parameter, the smallest value consistent with performance at the human level and with the other two parameters set accordingly.

The results of this simulation with the 7 different sets of parameters are shown in Figure 5. Red symbols indicate mean psychophysical performance. The top left panel shows the performance when all three parameters were set to the optimum. Each of the other 3 panels shows the effects of placing one of the parameters at extreme values that are just inside the 30° performance contours of Figure 4. As can be seen, all combinations of parameters match the data well, although the optimum set of parameters generates consistently better performance.

**Simulation of association fields: Discussion**

There is a broad agreement between the results obtained for this simple association field model and the psychophysical data. Both show a similar level of performance and a similar dependence on path angle and on path length. The most interesting finding is that there is little effect of the association field parameters, outside of some very broad boundary conditions. The curvature term can be set to as large as 175° and the co-circularity term can be set to 18°. In effect, each parameter simply provides a mechanism for applying more weight to Gabor element relationships that are closer to each other in orientation.
The values of the parameters set the gradient of this differential weighting by placing it on different parts of the Gaussian curve, but not much else and so consequently have little effect.

However, the distance term shows a more constrained requirement: it needs to be set to between $0.75\Delta$ and $2\Delta$. It is important for the association field to be neither too small nor too large with respect to the element spacing in the field.

It is important to note that this form of model does not identify or compute anything about cross element structures, such as contours. For example, the set of element weights that is produced does not provide any information about the length or the precise configuration of the target path. An element having a high total weight value does not know what contributions have led to that high value.

### Explicit linking and contour description

A different approach to the computational task of the path-detection paradigm is to explicitly link sets of elements and then to inspect the contour of the chains so formed. The process has these steps:

1. linking process: determine pairwise links between elements and concatenate linked pairs into chains of various lengths; and
2. contour process: measure the curvature properties of the chains and use these heuristically to identify the target.

### Forming chains

Across an entire stimulus field, the full set of pairwise element links are obtained according to the type of linking model being used. These pairwise linked elements can be concatenated into chains of any specified length. From the set of links, it is possible to generate the full set of chains of any particular length. The numbers involved tend to be enormous.

### Contour bends along a chain

Each chain is characterized as a contour with a sequence of bends: changes of contour orientation at each element along the chain. In common with the logic of the psychophysical paradigm, we suppose that the likelihood that any chain is the target is determined by how great the largest bend along the chain is. Consequently, the model finds the chain of lowest maximum bend. A response is generated depending on which side of the stimulus this contour is found.

A contour along a chain passes through each element center in turn so that the element is a local tangent. Since the contour can pass through each element in 2 different directions (with tangents that are strictly 180° apart), for a chain of length $m$ there are $2^m$ possible contours. Across each link between 2 successive elements in a chain, the

---

Figure 5. The effects of path length on performance of the various association fields. The top left panel shows the association field with all 3 parameters set to their optimum values. In each case, the red points are the human psychophysical data.
The implicit contour has 3 distinct directions: the direction of the first element, the direction of the link, and the direction of the second element. It follows that each link has 2 changes of contour direction (2 bends). Similarly at each element there are 2 bends, and the path angle is the sum of these bends.

Any one contour through a chain of \( m \) elements therefore comprises a sequence of \( 2m - 1 \) directions (\( m \) element directions plus \( m - 1 \) link directions) and therefore \( 2m - 2 \) changes of direction. We will represent the contour by a vector of \( 2m - 2 \) bend values. For a given chain, it is possible to enumerate the vector of bend values for each of the possible contours (arising for different element directions), giving \( 2^m \) vectors, each \( 2m - 2 \) in length. The best contour for this chain is then the one with the lowest largest bend. The largest bend in this best contour through a chain will be called the chain bend. The example in Figure 6 shows the chain with the smallest chain bend in the field (marked in yellow) and the link in that chain with the largest bend (marked in red).

Chain bend, the largest bend in a chain, provides a heuristic for deciding which chain is the target path in all of the simulations that follow. The target is identified as the chain with the lowest chain bend.

Model family 2: Isotropic element linking

The association field is both anisotropic in what elements are implicitly linked and selective for the orientation form that the implicit link makes. The second family of models we consider is different in both regards. Linking is isotropic and based only on position (“where”) not form (“what”).

We report simulations of 2 different isotropic linking processes. Each is based on the proposition that linked elements should be neighbors in some sense.

Linking process 1: Fixed distance

A conceptually simple model for linking neighbors is to use a fixed critical distance. Any pair of elements that are separated by less than the critical distance are linked, as illustrated in Figure 7.

Figure 6. A piece of a typical path stimulus showing (thin blue lines) the links between neighboring elements. The yellow lines indicate the chain of length 3 with the best bend (i.e., the chain for which the maximum bend is smallest). The bends at each change in contour direction are shown, with the largest bend in red.

Figure 7. Fixed distance links for a sample noise field. The left panel shows the links that are found over a distance of 0.8\( \Delta \) and the right panel shows links over a distance of 1.2\( \Delta \).
The first step was to run a series of simulations of the basic experiment, with targets of path length 8 and path angles between $0^\circ$ and $40^\circ$. The effects of path angle are similar to the psychophysical data: declining performance with increasing path angle. The path angle consistent with 75% correct performance is shown in Figure 8 as a function of the critical distance. As can be seen, performance reaches a peak for a distance of around $1.4\Delta$ and declines again thereafter. Distances from $1\Delta$ upwards are sufficient to generate performance that is at least equal to the human data for this condition (indicated by the horizontal line).

Next, a simulation of the main experiment was performed with a range of values for the critical distance. Maximum tolerable path angles are plotted as a function of path length in Figure 8, which also shows the mean psychophysical data for comparison. Functions for various different critical distance parameters are shown. The performance of this model is broadly similar to (and slightly better than) the human data, for all distances greater than $1\Delta$. At all distances, the dependence on path length is similar to that found in the human data.

As with previous models, the distance term has a non-monotonic effect on performance. Once again, there is a cost associated with a mechanism that is too large.

### Linking process 2: Delaunay triangulation

Fixed distance linking requires that the distance term be set to match the structure of the image. We now consider an isotropic linking process that is self-organizing in that respect.

The Delaunay triangulation offers an intuitive (but arbitrary) possibility for determining neighboring elements. The process finds all possible pairs of elements such that there exists some locus in the image that is equidistant from each member of the pair but that is further from all other elements.

A sample Delaunay triangulation is shown in Figure 9. Each element is a member of some number of triangles, usually between 5 and 8, providing it with that number of neighbors. The Delaunay triangulation has a minor but
very unhelpful instability. Four points that form a trapezium (2 triangles sharing a common edge) will lead to 5 links: 4 around the outside and 1 along the shorter diagonal. As the angle of the trapezium smoothly changes, the identity of the shortest diagonal discontinuously flips. So a tiny change in the configuration of the 4 points can lead to a substantial change in the pattern of adjacency.

An enhancement to Delaunay triangulation that reduces this instability substantially is to link the other pair of opposing corners for each trapezium. We add that extra connection (doubling the number of links across the image) and have a more robust adjacency measure. We refer to this triangulation as Delaunay++.

A simulation of the full experiment was conducted for both Delaunay and Delaunay++ triangulations. The results are plotted as a function of path length in Figure 10, which also shows the mean psychophysical data for comparison. Performance with the Delaunay++ triangulation is a little worse than the simple Delaunay because it generates more links. However, both are a little better than the human data and show a similar dependence on path length.

Note that Delaunay++, which can be characterized as a mechanism with a wider spatial extent, produces poorer performance. As with the association field, there is a cost to extending the linking process too far in space. Unlike the association field, however, the scale of the Delaunay process is set by the stimulus, not by a parameter of the model.

Model family 2: Discussion

The models explored in this section have the critical property that they link elements regardless of their orientations. The simulations show that the target path can be correctly identified, at least as well as the psychophysical data, by a least contour bend heuristic applied after linking.

The computational benefit of such models is that they open up the possibility of contours of other shapes being perceived. In these models the only constraint on contours is that the successive parts of a contour need to be adjacent in some sense. It is a useful result that the main psychophysical findings of the path-detection paradigm can be modelled in this way.

There is at least one finding that cannot be as readily accounted for by such models. There are large psychophysical effects of the orientation of an element with respect to the contour (Bex et al., 2001; Ledgeway et al., 2005). Elements oriented along the contour (0°) are most detectable; elements aligned at right angles (90°) to the contour next most, and elements aligned obliquely (45°) to the contour are essentially undetectable. The models described in this section link elements identically in the 3 cases, and given an appropriate (and different) heuristic for identifying each type of target, performance should be unaffected by the orientation of the elements to the contour axis. It is possible that variations in human performance reflect the availability or not of an appropriate heuristic for these other contour alignments. However, the next section explores the possibility that modifications to the linking process offer an alternative explanation.

Model family 3: Oriented linking

A completely non-oriented linking process is sufficient to model the human data on the basic contour-detection paradigm, provided links are restricted to adjacent elements. However, such models generate very high
numbers of links, many of which are unlikely to belong together. This section explores the effects of using a process of oriented links, which will reduce the number of links.

It is important to understand that the original association field concept is a “where” and “what” linking process. Two elements are linked if they are both oriented appropriately (i.e., both point to each other). In this section, we explore a form of just “where” linking that is oriented so that 2 neighboring elements are linked provided at least 1 of the elements points to the other. There is no selectivity for the “what” that is produced by the link.

**Linking process 3: Elliptical linking region**

A simple version of oriented linking is to use the fixed distance process, but with a linking region that is elliptical rather than circular and where the orientation of the major axis of the ellipse is aligned with the orientation of the element. Although many other forms of linking region could be devised, this is sufficient to illustrate the effect of an oriented linking region.

The effects of varying the aspect ratio of the elliptical linking region are shown in Figure 11. Each panel shows data for a different value of the ellipse major radius (1.25, 1.5, 1.75). In each panel, each curve describes the effect of path length on maximum tolerable path angle for a different aspect ratio, varying from 0.14 (lowest curve) to 1 (highest curve). The data show that there are many combinations of major and minor axis radii that can produce performance equal to or better than human performance.

**Linking process 4: Oriented linear filter response**

A more complex model for oriented adjacency is to use zero-bounded regions in filtered images. Zero-bounded regions of filter response have been used by Watt and Morgan (1985) in a theory of the primitive spatial code, limited to 1-dimensional considerations, and a 2-dimensional treatment can be found in Watt (1988, 1991). They have also been used by Dakin and Hess (1998) to explore contour linking.

For the current study, we have used filters with a form very similar to the Gabors themselves and are either elongated second derivatives of Gaussians, Gabors, or log Gabors. All of these filters are spatial frequency band-pass in one axis, with a bandwidth of around 1.5 octaves, and low-pass in the orthogonal axis. Their orientation bandwidth is of the order of 40°. As expected, the results for the different filter forms in all cases are essentially indistinguishable. Importantly the responses of filters are not limited to the Gabor elements themselves but tend to stray beyond them. A filter that has a spatial scale that is somewhat larger than the individual elements can produce a response that extends over 2 or more elements, provided they are aligned appropriately in space.

Any pair of elements that lie inside the same zero-bounded region are linked, as depicted in Figure 12.

![Figure 11. Performance for various ellipse linking regions. The 3 panels correspond to major axis radii of 1.25, 1.5, and 1.75. In each panel the different data are for different aspect ratios from 0.14 to 1.0. The smaller aspect ratios are the lower curves.](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/933536/ on 11/22/2018)
This model has 2 basic parameters: filter spatial scale and threshold. The first of these affects the amount of linking in a simple fashion: the larger the scale, the more elements that will be linked together. However, at the same time, the larger the filter scale, the lower the response amplitude and this will tend to limit the usefulness of larger filters. Filters with a spatial scale (expressed as a wavelength) of $2\Delta$ or more will produce a response to the stimulus that is less than 5% of the mean unsigned response of the same filters to natural images. Such filters are considered to be of no value.

The main computational purpose of the threshold value is to remove suboptimal responses. The threshold value also affects the amount of linking: the higher the threshold, the fewer elements that will be linked together. Thresholds are expressed as a proportion of the maximum response of the filter to the stimulus.

The joint effects of filter scale and threshold value on performance were assessed. An experiment with a target path length of 8 elements and path angles between 0° and 40° was simulated. For each combination of filter scale and threshold, the path angle consistent with 75% correct performance (i.e., the maximum tolerable path angle) was measured. This is shown in Figure 13. The black line encloses the combinations of filter scale and threshold that result in performance at least as good as the human data. As can be seen, there is quite a wide range of spatial scales (at least an octave) and threshold values that generate performance as good as the human data for a path length of 8 elements.

The lower panel in Figure 13 shows the effect of spatial scale, for a single threshold value of 0.35. As can be seen the effect is strongly non-monotonic, with a peak at a spatial scale of around $1\Delta$.

In the next step, a full simulation of the main experiment with target path lengths varying from 3 to 8 elements was performed for 4 combinations of threshold value and spatial scale. The combinations chosen were

i. the optimum in Figure 13;
ii. the optimum threshold with the largest scale consistent with human performance;
iii. the optimum threshold with the smallest scale consistent with human performance; and
iv. the optimum scale and the largest threshold consistent with human performance.

The maximum tolerable path angle as a function of path length for each combination of parameters is shown in Figure 14.

Model family 3: Discussion

In common with earlier models, the oriented linking family of models behaves in a similar fashion to the human data, but slightly better. Once again, there is a cost associated with using a mechanism that is too large.

These models have the feature that they will not link Gabor elements that are oriented obliquely to the contour axis. Within the framework offered by such models it would be the case that targets of this type cannot be detected in human vision (e.g., Ledgeway et al., 2005) because the constituent elements along the path are not linked to each other and are probably linked to adjacent background elements.

General discussion of results

The principal aim of the present study was to explore the computational stages in solving the psychophysical
path-detection task of Field et al. (1993) by comparing the behavior of qualitatively different families of models. All of the models explored showed the same basic effects of path angle and path length and revealed a fairly close requirement on the spatial size of the mechanism involved.

The first family of models considered was an implementation of the association field of Field et al. (1993). In this family, the target path is detected when 1 or more of its elements have the highest overall association strength in the image. Association strength arises when an element is surrounded by other elements with positions and orientations that form smooth and, ideally, co-circular contours.

The remaining 2 families of models were each based upon an adjacency constraint to make explicit the link between neighboring elements. Links are concatenated into chains of elements. These models produce a large number of chains of elements for each stimulus, and the chain that has lowest curvature is selected. When this chain is the target path, then the model has performed correctly. Unlike the association field, the second family of models did not require any “what” selectivity for the orientation relations between elements in the image at the linking stage. Interestingly despite being different, the optimal version of models in this family exhibited a very similar pattern of performance to the human observers and the association field.

The third family of models restricted links between elements to cases where 1 of the linked elements was located in the direction of the orientation of the other element but could be of any orientation itself (i.e., “where” but not “what”). These models also showed the same pattern of behavior.

Before reaching some general concluding remarks, the two main generic findings will be considered in more detail.
Path angle

Figure 15 shows that the various models are difficult to distinguish on the basis of their performance as a function of path angle and path length (when the optimal set of parameters is used for each). This suggests that the maximum path angle that can be detected is largely determined by the statistics of the randomly oriented background elements.

To check whether image statistics could be limiting performance in all cases, we have measured the distribution of values of smallest chain bend in noise fields, composed of randomly positioned and oriented elements. Chains were created using Delaunay triangulation to identify adjacent elements. All possible chains of a given length were then enumerated and the best, the one with smallest chain bend, identified. Figure 15 shows the median smallest chain bend as a function of chain length for 250 noise fields. The figure shows that for a chain of length 8, 50% of noise fields will contain, by chance, a chain bend of 38° or less. Note that under our terminology, a target path angle comprises 2 nearly equal bends. A target of length 8 elements and path angle 70° will have a maximum chain bend of around 38°. So the data are comfortably above the psychophysical data but do have the same form. The presence of spurious low curvature contours in the noise field will have a significant impact on target detectability, and these results place an upper limit on human performance. It is important to note that the actual values in these data, in the psychophysics, and

Figure 15. Performance of each model, expressed as maximum tolerable path angle, at its optimum parameter values, compared with the human data (left panel). The different models exhibit similar performance and all are somewhat better than the human data. The right panel shows the median best chain bend per image for noise fields alone.
in the computational results are all dependent on the size of the field. In the limit, an infinite noise field will have an infinitely long perfectly straight contour.

As well as illuminating the path angle results of the path-detection paradigm, the data above provide a useful general statistic. Natural images have a pattern of contour that is mostly random in the sense that it arises from unrelated objects. It is plausible that these can be modelled to a first approximation by the noise fields we have used: fully packed fields of randomly located and oriented elements subject to a minimum separation constraint. If so, then the data of this section will theoretically describe the accidental aspects of natural images.

Spatial size of mechanism

An important property of all of the models investigated in the current study is that they each show a dependence on the spatial size of the purported linking mechanism. The spatial parameters used to describe the size of a mechanism vary from model to model but, in each case, it is possible to convert the size parameter to a measure of the maximum distance over which 2 elements will be linked. As throughout, the unit for distance is the minimum inter-element separation. When this is done, and the variations in performance are plotted as a function of the maximum distance for the various models, a common function emerges. This is shown in Figure 16. As can be seen, provided the mechanism is large enough to link elements together (max distance greater than 1Δ) but sufficiently small so as not to link elements that are further apart than the maximum inter-element separation (2.5Δ), then performance is equal to or better than the human data. Thus, for optimal performance the spatial structure of each model needs to be tuned to the density of the elements in the visual field. In achieving this, some models (such as the Delaunay triangulation model and the linear oriented filter response model) scale naturally to the stimulus characteristics, whereas the others require hard-wired parameters to take into account the stimulus properties.

The adverse consequences of using linking distances that are too large are themselves a consequence of the curvature structure of the noise field. When links occur over larger distances more links will be made, and exponentially more chains can be created. The chain bend of the straightest spurious chain is lower when more chains are created (this is a property of the Generalised Extreme Value Distribution Type III—Weibull). Consequently, a target has itself to be straighter to be reliably detected.

The finding of a psychophysical density limit, such that paths cannot be readily detected at inter-element separations greater than 4 to 6 times the Gabor wavelength (e.g., Kovács & Julesz, 1993), might suggest a limitation on the degree to which the spatial structure of the mechanism can be determined by the stimulus structure. Importantly the linear oriented response filter model readily exhibits a natural limitation of this type since the maximum size of the filter is determined by the Gabor wavelength (because filters with a size greater than about 4 times this will respond too weakly to be of value). The largest filter and the smallest reasonable threshold together would fail to link elements, even when appropriately aligned in orientation, if those elements were more than about 5 wavelengths apart.
General remarks

This paper has several outcomes. First, we have established that some of the psychophysical findings concerning the contour-detection paradigm of Field et al. (1993) could in principle be due to the structure of the stimulus rather than limitations imposed by the human visual system. For example, we have found that, even with the association field family of models, there is no need to tailor the orientation and curvature related terms to the detectable stimuli. On the basis of data from this paradigm, there is no evidence for a visual system limitation to bends of 40° or thereabouts, although there may be other data that lead to this conclusion (e.g., see Watt & Andrews, 1982).

The second outcome is that we have demonstrated the importance of element separation as a determinant of linking. This is consistent with other approaches (e.g., Elder & Goldberg, 2002; Elder, Morgenstern, & Gabone, 2003; Kubovy, Holcombe, & Wagemans, 1998).

The third outcome is that we have explored the usefulness of separating contour linking processes from contour characterizing processes. In the bare form that we have used here, the original association field achieves neither. When an element has a high degree of association strength, there is no information about which of the elements around it are responsible, so it has not achieved an explicit linking. Equally, in common with all templates, this scalar value is a product of amount of contour and quality of contour and so cues neither. It would be relatively straightforward to model a version of the association field that made explicit links between elements, and we could have studied a fourth family of models with linking restricted both spatially and bi-orientationally. There was no need because the families we have used are more general than that.

Previous research has shown that performance on the path-detection paradigm is sensitive to the orientation of the Gabor elements with respect to the orientation of the contour they form (e.g., Bex et al., 2001; Field et al., 1993; Ledgeway et al., 2005). When the elements are oriented to be orthogonal (“ladders”) to the contour reliable detection is possible, but performance is more affected by path angle than when the elements are aligned with the contour (“snakes”). Ledgeway et al. (2005) found that stimuli in which the elements are all aligned at 45° to the path (“saw tooth”) are essentially undetectable. There is no single and completely satisfactory account of these findings, including among the models discussed here. However, our formulation of the phenomenon can be used to explore the issue further.

When targets can be detected, we would argue that 3 conditions must be met:

i. most or all of the target elements must be linked together;

ii. a description of the shape of the resultant contour must be created; and

iii. there must be some quality of that shape description that is sufficient to distinguish the target from spurious “noise” contours.

When targets are not detected, then one of these conditions has failed.

In this paper we have considered only one option for the second condition: contour shape described by largest bend along the contour. For “snake” targets, that in combination with any form of linking is enough to explain the pattern of performance. Furthermore we have shown that performance fails when the third condition is not met.

For “ladder” targets with small detectable path angles all 3 conditions will be met. This implies either isotropic linking or some other form of oriented linking that we have not considered in this paper. The same targets with larger path angles are not detected. This cannot be accounted for by condition (iii) noise field statistics since these are the same as for the “snakes” case. There are various possible ways to explain this, such as spurious competing collinear links between target elements and adjacent noise elements.

The inability to detect contours when the constituent elements are oriented obliquely with respect to that contour (“saw tooth”) could be due to an inability to link the elements of the target, or a similar effect of competing collinear links with the noise elements. Alternatively it could indicate successful isotropic linking followed by an inability to produce a description of the contour bends that allows it to be adequately distinguished from other contours.

These interesting cases raise questions that are beyond the scope of the present work. If there are multiple ways of linking, then do different forms of linking compete, or can the same element simultaneously be linked in different ways? What other contour descriptions might be useful?

Acknowledgments

SCD acknowledges the support of the Wellcome Trust.

Commercial relationships: none.
Corresponding author: Roger J. Watt.
Email: r.j.watt@stirling.ac.uk.
Address: Department of Psychology, University of Stirling, Stirling, Scotland, Fk9, 4LA.

References


