International Journal of Computer Vision 55(1), 5–25, 2003

Perceptual Grouping by Selection of a Logically Minimal Model

JACOB FELDMAN
Department of Psychology, Center for Cognitive Science, Rutgers University, New Brunswick, NJ 08903, USA
jacob@ruccs.rutgers.edu

Received March 9, 2000; Revised February 14, 2002; Accepted March 26, 2003

Abstract. This paper presents a logic-based approach to grouping and perceptual organization, called Minimal Model theory, and presents efficient methods for computing interpretations in this framework. Grouping interpretations are first defined as logical structures, built out of atomic qualitative scene descriptors ("regularities") that are derived from considerations of non-accidentalness. These interpretations can then be partially ordered by their degree of regularity or constraint (measured numerically by their *logical depth*). The Genericity Constraint—the principle that interpretations should minimize coincidences in the observed configuration—dictates that the preferred interpretation will be the minimum in this partial order, i.e. the interpretation with *maximum depth*. This maximum-depth interpretation, also called the *minimal model* or *minimal interpretation*, is in a sense the "simplest" (algebraically minimal) interpretation available of the image configuration. As a side-effect, the "most salient" or most structured part of the scene can be identified, as the maximum-depth subtree of the minimal model. An efficient ($O(n^2)$) method for computing the minimal interpretation is presented, along with examples. Computational experiments show that the algorithm performs well under a wide range of parameter settings.

Keywords: perceptual grouping, perceptual organization, logic, nonaccidental properties

This paper summarizes a novel logic-based framework for grouping and perceptual organization, and introduces an algorithm for computing interpretations in it efficiently. The framework was originally conceived as an abstract account of the human perceptual organization faculty. Grouping and perceptual organization have proven to be remarkably difficult problems, and human observers' subjective judgments are still far superior to the best known computational techniques. Hence a formal account of human intuitions is a potentially crucial step towards the development of better computational techniques, insofar as the account is concrete enough to lead to computable procedures. Formal and theoretical aspects of this framework are discussed more completely elsewhere (Feldman, 1997b). This paper summarizes the framework informally, in enough detail to motivate the algorithm, and also applies the framework to the extraction of salient structure from images.

The approach described in this paper differs in several ways from existing algorithms for grouping, in part reflecting its origins as an attempt to render human perceptual organization computationally. Before describing the theory and algorithm in detail, it is worth explaining and defending some of the most salient differences from existing approaches.

Hierarchical Representation

In many computational grouping theories, the goal of grouping is assumed to be a partition of the image elements into equivalence classes, which are hoped to correspond to genuinely distinct objects. By contrast, research in human perception has generally found that the visual system organizes the image *hierarchically*, that is, using a more complex multi-scale description with more abstract relations represented at

higher levels. (Hierarchical representations are also produced by a number of existing computational approaches [e.g. Shi and Malik, 2000]; these will be discussed more below). A number of influential papers in the psychological literature (Baylis and Driver, 1993; Palmer, 1977; Pomerantz et al., 1997) have argued that representations of image structure are hierarchical, simultaneously containing descriptions of local structure and superordinate global structures on multiple levels. More recently others (Markman and Gentner, 1993; Medin et al., 1990; Palmer, 1978) have shown that judgments of visual similarity among images depend on these hierarchical relations, corroborating the idea that they are central to perceptual descriptions. Moreover, Pomerantz and Pristach (1989) have argued that grouping itself is a side-effect of this hierarchical description; rather than visual items simply being aggregated, items may adhere together only when they serve together as arguments to a common descriptive predicate in the hierarchy. Under this view, grouping per se is secondary to a primary process of rich scene description.

A key element in the visual system's hierarchical representation is the existence of relatively abstract organizational principles operating at the level of large image regions and structures, alongside relatively low-level principles operating at the level of individual image elements. Although the computational grouping literature contains frequent allusions to Gestalt grouping principles, discussion is usually limited to the principles of proximity, "good continuation" (collinearity), and similarity, each of which operates at the level of individual (oriented) elements. The original Gestaltists' lists of principles were far longer and more comprehensive (comprising as many as 114 distinct principles, according to Pomerantz (1986)). Many of the principles usually omitted from mention in the computational literature involve more abstract relations that cannot be computed among simple image elements. Among these, some that have received modern empirical support include *symmetry* (Kanizsa, 1979; Sekuler et al., 1994; Wagemans, 1993; Wagemans et al., 1993), *closure* (Elder and Zucker, 1993; Kovacs and Julesz, 1993), and *convexity* (Palmer, 1992; see also Jacobs, 1996 for a computational treatment).

Qualitative Perceptual Interpretations

Another salient and potentially controversial aspect of the current research lies in its reliance on *qualitative assertions* or predicates about image structure, rather than quantitative estimates of scene parameters. In the theory presented below, scene interpretation constitutes a choice among qualitatively distinct interpretations or models. The main goals of the theory are (a) to specify the appropriate space of distinct possible interpretations, and (b) to articulate a preference rule by which one should be chosen. Again, the motivation for this stems from what is known about human perceptual organization.

First, there is a long tradition in the study of human vision of regarding perception as a choice among qualitatively distinct alternative interpretations, and a great deal of evidence for this view (Gilchrist and Jacobsen, 1989). Rubin's famous vase/face figure, reproduced in many psychology textbooks, is an illustration that subjective perceptual interpretations can take on multiple distinct and mutually inconsistent forms. Each interpretation (vase or faces) is qualitatively distinct from the other, and produces a completely different subjective percept. In multi-stable examples like this, the human observer tends to alternate between distinct percepts. But the point extends more generally to normal, stable percepts. The "unconscious inference" yielding a single winning percept in normal perception seems to be the result of a process by which this winning percept defeats a set of qualitatively distinct alternatives (of which the observer may never be consciously aware).

Figure 1 gives several other famous examples of choice among qualitatively distinct interpretations. Figure 1(a) shows Attneave's (1968) triangles, which were studied more extensively later by Palmer (1980). Equilateral triangles can be seen as "pointing" (bearing a primary axis) in any of three distinct possible directions (Fig. 1(a)). When several triangles are aligned (Fig. 1(b)), they are all perceived as pointing along the common main axis of the set. In the latter case the single "winning" orientation is the result of a choice among (at least) the three distinct alternatives. Perceptual completion (Fig. 1(c)) is another example of a problem where the system seems to see either one alternative (one square occluding another) or a completely distinct alternative (a square abutting an L-shaped object). Boselie and Wouterlood (1989), Sekuler and Palmer (1992) and van Lier et al. (1995) among others have recently investigated the factors that lead human observers to alternate among the possible interpretations. Another example is provided by the perception of regular grids of dots, called dot lattices (Fig. 1(d)), investigated by Kubovy (1994), Kubovy

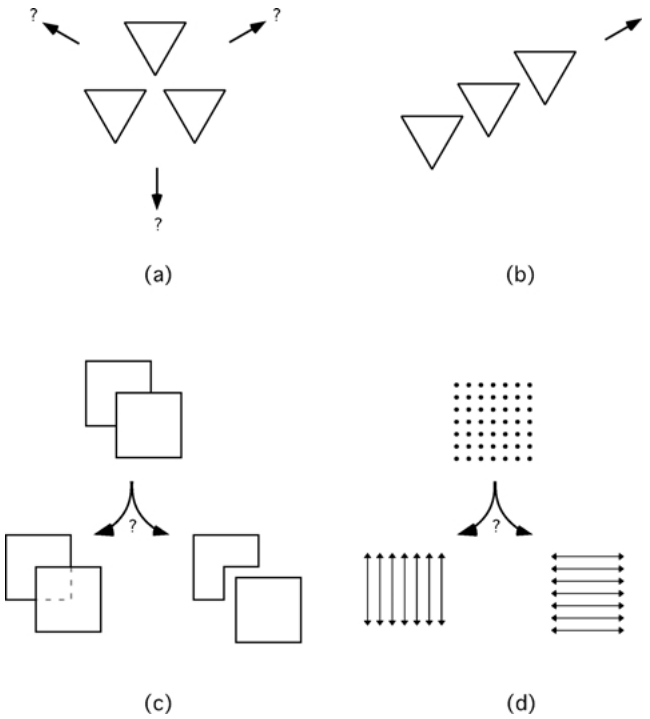


Figure 1. Examples of situations where the human perceptual system seems to choose among qualitatively distinct alternative interpretations. (a) *Ambiguous triangles* (after Attneave, 1968). The three triangles can be seen as "pointing" in any of three directions (b) When aligned, the triangles' orientation no longer seems ambiguous. (c) *Completion behind an occluder.* Does one see a square behind a square (left) or a square abutting an L-shaped object? (d) *Perception of dot grids.* The regular grid of dots can be seen as vertically or horizontally aligned.

and Wagemans (1995) and Zucker et al. (1983). These may be interpreted either as aligned horizontally or aligned vertically, but, typically, not both. Small changes in the spacing or luminance of rows can cause abrupt alternations between the vertical and horizontal interpretations.

The prominence of qualitative assertions in the theory below is not, however, intended to imply that quantitative estimation does not play an important role in grouping. The current author has in fact provided evidence that Bayesian theory provides a good quantitative account of human subjects' judgments in a grouping task (Feldman, 2001b). The point rather is that the *space of possible interpretations* is qualitative. Bayesian theory does not in and of itself provide the interpretation space, but only attaches quantitative strengths to each interpretation given the data. Hence the theory below emphasizes the construction of the appropriate space, and gives a simple qualitative rule for selecting an interpretation from this space. In close cases a more precise quantitative rule might be required, but such a computation is far more expensive, and the examples and discussion below suggest is often unnecessary. Indeed, the qualitative rule discussed below can be shown to be formally equivalent to Bayes' rule under certain highly simplified assumptions (equal priors and truncated/thresholded likelihood functions; see Feldman, 2001a for discussion).

It is worth noting that, aside from its resemblance to human perception, the production of qualitative predicates about the image is computationally useful. Many practical decisions the agent must face are inherently qualitative. (Should a particular part of the image be matched to a database or not? Should a limb be moved? Which limb? Fight or flight?) In turn many of these decisions depend on qualitative categories of image structure. (Food or rock? Friend or foe?) At some point, quantitative parametric image descriptions must be converted into such "meaningful" qualitative categories in order to support action.

We now turn to a more specific description of the theory underlying the algorithm.

The Logical Approach

In both computational and psychological theories of visual interpretation, it is commonplace to speak of constructing a "model" of the observed scene. As discussed above, such models are often qualitative in nature, with one model of a given scene being distinctly and sharply distinguishable from alternative models, as discussed above. The term "model" is highly suggestive of mathematical logic, where the construction of formal models is of primary concern. Indeed, Reiter and Mackworth (1989) have proposed bringing the full machinery of mathematical logic to bear on visual interpretation problems, as supplemented by modern methods from Logic Programming for performing efficient computations on logical expressions. More recently others have integrated logical operators (Iverson and Zucker, 1995) or "qualitative" probabilities (Jepson and Mann, 1999) into conventional continuous methods of image analysis with impressive results. In principle the logical approach has the tremendous advantage of affording an exhaustive enumeration of the space of interpretations that are possible for a given scene, allowing one to (a) characterize the internal structure of this space and (b) identify a criterion for choosing a "best" interpretation from it. This in turn means that one can begin to explore the formal semantics of visual inference—e.g., give explicit truth conditions for visual interpretations in the style of mathematical logic. Perhaps due to the lack of an explicit logical model of visual interpretation, this seemingly natural goal of vision science has not as yet been pursued.

Imagine that one had a way of enumerating all the models consistent with a given image. In order to construct a full interpretation theory, one would need in addition some way of inducing a *preference ranking* on these models, so that superior models may be chosen over inferior ones. This leads to the idea of a *partial order* among models, the explicit construction of which forms the core of MM theory. But what is the basis for preference among models? That is, when one chooses a visual interpretation among the many that are consistent with the image, what is one trying to optimize?

Genericity

Once answer comes from the *Genericity Constraint*, i.e. "avoid accidental configurations," or, equivalently, choose the model in which the observed configuration is most typical. This idea originated in 3-D reconstruction, where it took the form of the generic *viewpoint* constraint (see e.g. Freeman, 1994; Takeichi et al., 1995). But the principle extends beyond viewpoint, to any scene parameter along which an unusual or special value signals an unsound perceptual inference (see e.g. Bennett et al., 1989; Kitazaki and Shimojo, 1996; Nakayama and Shimojo, 1992). Non-accidental properties (Binford, 1981; Lowe, 1987; Witkin and Tenenbaum, 1983) illustrate the logic. If two line segments (in the image) originated nearby each other along the same smooth object contour (in the world), then they will be collinear generically (that is, typically); whereas if they originated independently, then any collinearity is accidental. Hence collinearity is a reliable cue that two collinear segments should be grouped together (Feldman, 1996; Jepson and Richards, 1992). Other properties can be chosen that, like collinearity, pick out a configurations that are subspaces of a generic space of configurations (just as non-generic viewpoints form a 1-parameter curve embedded in the 2-D viewsphere, and collinear line segments form a point in the 1-parameter angle space).

The approach taken below is to stipulate a fixed set of such non-generic configurations, which are called *regularities*. An observed configuration can then be characterized by the set of regularities it obeys. A *model*, then, is simply a set of regularities that a configuration obeys, which characterizes the configuration qualitatively. A partial order can then be induced on these models by subset inclusion in the regularity set. That is, a model M_1 is strictly more regular than M_2 if M_1 exhibits all the regularities that M_2 exhibits, plus at least one more. This partial order in effect ranks the models by their degree of "coincidentalness," which can be measured numerically by the number of regularities in the model, called the *logical depth* or *codimension*[1] (Jepson and Richards, 1991). Following the Genericity Constraint, the most preferred model is then the one with *maximum depth*. This optimization rule, it will be shown below, leads directly to a procedure for efficiently choosing the intuitively preferable grouping interpretation.

Minimal Model Theory

Consider the problem of grouping a field of localized visual items (e.g., dots) in the plane (Fig. 2). Much research on this problem has appeared, ranging from classic Gestalt principles, psychophysical studies and neural accounts (Compton and Logan, 1993; Glass, 1969; Kubovy et al., 1998; Prazdny, 1984; Smits and Vos,

Figure 2. A field of dots, in which the eye naturally finds a cluster and a chain.

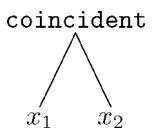
1987; Zucker et al., 1983) to computational schemes (Guy and Medioni, 1996; Parent and Zucker, 1989; Stevens, 1978; Stevens and Brookes, 1987; Zucker, 1985). Minimal Model (MM) theory is based on the idea of grouping together those items whose geometrical configuration is unlikely to be an accident, and hence whose locations are unlikely to have arisen independently.

In selecting a regularity set, one begins by identifying atomic configurations that are unlikely to occur "by accident." One obvious choice is *coincidence.* Assume first that dot generating processes are spatially localized. From this simple assumption it follows that for two dots $x_1, x_2 \in R^2$, the predicate

$$\texttt{coincident}\,(x_1, x_2) \leftarrow \|x_1 x_2\| < \theta_{\text{coincident}} \qquad (1)$$

is a regularity, in that it will typically hold (i.e., is generic) if the two dots originated from the same localized generating process, but *non*generic if they did not (in which case their locations ought to be independent). In the expression, $\|x_1 x_2\|$ denotes the distance in the plane between the two dots, $\theta_{\text{coincident}}$ is some distance threshold, and "$\leftarrow$" should be read as "if."[2] Admittedly, the use of thresholds (rather than smooth affinity functions) is probably not a psychologically plausible definition, in part because it is not scale-invariant. Other superior definitions might be adopted, but the examples below suggest that this definition works surprisingly well. In any case the use of threshold is similar to the common practice of thresholding affinities to yield an adjacency matrix (e.g., Kelly and Hancock, 2000).

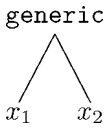
It is convenient to express a coincidence (or indeed any other predicate) between two dots x_1 and x_2 as a tree:



(2)

In general, the head term of a tree is a model M consisting of a set of predicates that describe the relationship among its subtrees (the arguments of the predicates). Hence the number of subtrees reflects the arity of the predicates in M. The size $|M|$ of the model, denoted depth(M), indicates how "regular" the relationship is: the more terms it contains, the tighter the relationship among its arguments (subtrees).

Arguments (subtrees) that make the head term "true" are said to *satisfy* it; e.g. two coincident dots x_1 and x_2 satisfy tree (2). By contrast, if the two are not coincident—and indeed have no other regular relationship recognized in the stipulated regularity set—then they are said to have a `generic`, i.e. "not regular" relationship:



(3)

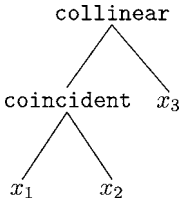
Here `generic` is an alias for the empty model $\emptyset$. Note that the `generic` relationship is *strictly weaker* than any regular relationship. Any configuration that satisfies a given tree also satisfies any strictly weaker tree. Hence two coincident dots satisfy both tree (2) and tree (3)—but they do not satisfy tree (3) *generically,* simply because they *also* satisfy the stronger tree (2). That is, a configuration satisfies a model generically when it does *not* satisfy any strictly stronger model[3](Feldman, 1997c). Hence defining the partial order—which explicitly ranks the models by relative strength—is crucial to defining genericity formally.

Returning to the enumeration of the regularity set, assume that there exist curvilinear generating processes in the world (i.e. processes localized in one dimension rather than two) (Feldman, 1997a; Parent and Zucker, 1989; Pizlo et al., 1997; Smits and Vos, 1986; Zucker, 1985). Hence, a second regularity is *collinearity:*

$$\texttt{collinear}(x_1, x_2, x_3) \leftarrow (\pi - \angle x_1 x_2 x_3) < \theta_{\text{collinear}}, \qquad (4)$$

where $\theta_{\text{collinear}}$ is some angle threshold (the disclaimers concerning thresholds mentioned above apply here as well). This relation is generic among three dots if they were generated by a common curvilinear process, but non-generic if they were generated independently. Because two coincident dots define an orientation,[4]this regularity can be expressed conveniently in tree form as a `collinear` relation between one dot x_3 and a

coincident subtree spanning x_1 and x_2, i.e.



(5)

(See the discussion of parameter passing below.) It is further stipulated that collinearity is only computed between items that are coincident, not between distant items; i.e.,

$$\texttt{collinear} \rightarrow \texttt{coincident}. \qquad (6)$$

The recursive embedding of trees illustrated in tree (5) can be used to construct arbitrarily large trees. Such a tree, called a *parse tree*, coupled with an assignment of each of its leaves to the members of some set $x_1 \ldots x_n$ of dots, constitutes a kind of qualitative interpretation of the dot configuration. Each node of a parse tree is a model of the relationship between its children, whether they are subtrees or leaves (i.e. dots). With a fixed finite set of regularities, there is only a finite number of models to choose from. With the regularity set stipulated so far, there are exactly three possible models: $\emptyset$, {coincident} and {coincident, collinear} (abbreviated generic, coincident, and collinear, or sometimes gen, coinc, and coll, respectively). These three models are partially ordered by their degree of regularity—i.e. the number of regularities they contain. In this case, the partial order is a linear chain (Fig. 3 shows the Hasse diagram). In general, with arbitrary regularity sets and pairwise implicational constraints, it is a distributive lattice.[5] Each model in the lattice has a unique depth (shown at the right in the figure), namely the number of regularities it contains; or, equivalently, the row number counting from 0 at the top row. Again, it should be emphasized that the choice of regularity sets if flexible. The set {coincident, collinear} with collinear → coincident has been used here for expository purposes, and will be used in the examples given below; but other more psychologically plausible sets may easily be adopted.

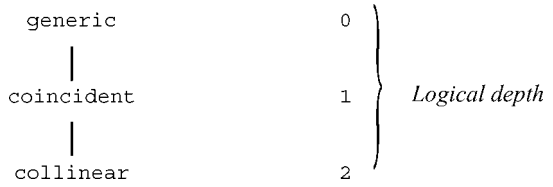


Figure 3. The lattice of models for the dots regularity set, partially ordering the models by their degree of regularity. Each model is a possible relationship between two subtrees. The depth of each model is indicated on the right.

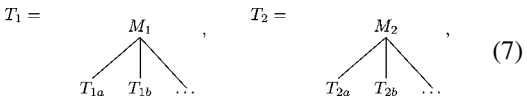
Rules for Passing Parameters

In order for a tree T to serve as an argument to some superordinate tree, it is necessary for T to pass parameters from the image elements in its scope to the superordinate tree. This parameter passing observes simple rules that depend on the type of tree, that is, the identity of T's head term. For example, if T is coincident, then it passes the location of the centroid of the elements in its scope. This has the same form as a single dot (i.e., spatial coordinates only), so T can now serve as an argument to any head term that takes dots as arguments. Similarly, a collinear chain passes orientations, namely the first and last orientations in the chain (oriented outward, i.e. in the direction of the chain's ends[6]), enabling superordinate trees to attach new elements only to the ends. In general, a head term of the form co-X would pass information of the form X. This rule is justified by the idea, underlying the entire theory, that if such a tree is the minimal model of the subtree, then the property X is the stable non-accidental regularity of the cluster. The one exception to this pattern is the notational convenience of having a coincident pair of dots pass the orientation defined by two locations, which then acts to "seed" a new collinear chain (see above).

A Preference Ranking

Crucially, the lattice of models is in fact a preference ranking, because of the Genericity Constraint: given that a certain regularity actually obtains in the image, a model that recognizes the regularity is always preferable to one that does not (all else being equal). The partial order can thus be used to induce a partial order on parse trees, by recursing on the tree structure. Because any tree has a model as its root node and trees as its arguments (which may of course be trivial, i.e.

leaves), we can denote two trees T_1 and T_2 by



(7)

where M_1 and M_2 are models chosen from the above list. Then we define the partial order recursively by $T_1 \leq T_2$ iff

1. $M_1 \leq M_2$, and
2. $T_{1i} \leq T_{2i}$ for $i = a, b, \ldots,$

where the partial order on models is defined as in Fig. 3. That is, the tree T_1 is at least as regular as the tree T_2 if T_1's head term is at least as regular as T_2's, and each of T_1's subtrees is at least as regular as each of T_2's corresponding subtrees. T_1 is minimally more regular than T_2—one "notch" more regular—if T_1's subtrees are exactly the same as T_2's subtrees, but T_1's head term is minimally more regular than T_2's head term; or, if T_1 and T_2 have the same head term, but exactly one of T_1's subtrees is minimally more regular than one of T_2's subtrees, and the rest are the same.

The recursive partial order naturally comes with a recursive version of logical depth. For a tree T with head term M and subtrees $T_a, T_b, \ldots,$

$$\text{depth}(T) = \text{depth}(M) + \sum_{i=a,b,\ldots} \text{depth}(T_i) \quad (8)$$

Leaf nodes (i.e. dots or other primitive elements) have depth 0 by definition. Thus a parse tree's depth is the sum of its head term's depth plus the recursive sum of all of its subtree's depths, and it increases in depth when any of them increase in depth.

Given an observed configuration of dots, one would like an interpretation that is "as generic as possible"—i.e., that leaves as few coincidences as possible unexplained (cf. Rock, 1983). This somewhat vague principle can now be restated as a concrete, computable rule, the *maximum-depth* rule:

(Maximum-depth rule)

Of all parse trees that dots $x_1 \ldots x_n$ satisfy, choose the one with maximum logical depth.

The maximum-depth interpretation is called the *minimal model* or *minimal interpretation*. Examples are shown below. In each case the interpretation is perceptually compelling. Not only is the set of dots partitioned in an intuitive way, but each cluster is characterized in a way that is qualitatively correct: as a curvilinear chain of dots, an unordered clump, etc.—or as having no particular structure whatsoever, i.e. "generic." The structure of the parse tree maps closely onto how one would verbally describe the configuration. Partly, of course, this reflects the psychological aptness of the regularity set chosen. One can easily imagine alternative regularity sets, involving such concepts as smoothness, cocircularity, etc., that are absent in the above regularity set. The main point is that whatever regularity set is chosen, the maximum-depth parse would be the preferred interpretation modulo that set. Expressing this fact more rigorously requires some discussion of the formal semantics, which is out of the scope of this paper (but again see Feldman, 1997b). It is important to note that the structure of the inference theory does not depend on the exact choice of regularity definitions. A different choice of regularity set simply leads to a different partial order, and thereby to different interpretations—but with analogous preference logic.

Moreover, the framework can easily be adapted to types of image primitives other than dots, e.g. line segments or edge elements. Though the leaves would then change their meaning, the logical structure of the interpretation theory would be unchanged. To underscore this point, examples are given below with a regularity set operating on edge elements instead of dots. For these examples, the regularity definitions need to be slightly modified (e.g. `collinear` can take two edge elements as arguments, but not two dots).

One theoretical advantage of the logical approach is that the partial order on parse trees enumerates *all* possible interpretations for a given family of scenes, e.g. for n dots (as always, modulo a particular choice of regularity set). The partial order thus amounts to a full interpretation space, in which the maximally preferred global interpretation is at the bottom and the least preferred is at the top. Figures 4–6 shows the full interpretation spaces for $n = 2$, $n = 3$ and $n = 4$ dots using the regularity set given above.

The size of the interpretation space grows exponentially with n, and for $n > 4$ the space becomes too large to display conveniently. For $n = 2$ and $n = 3$ the partial orders are linear chains, but for $n = 4$ and above, the structure becomes more complex. In general, the space is a set of disjoint distributive lattices[7](again see Feldman, 1997b for proof). This fact is pictorially evident in Fig. 6, in which the partial order consists of two disjoint parts.

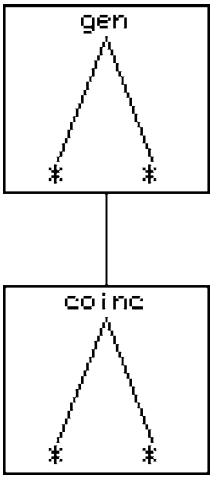


Figure 4. The interpretation space for $n = 2$ dots. Asterisks denote leaves (dots).

In practice, one would not normally actually construct the interpretation space for realistic values of n. Rather the interest in exhibiting and examining these spaces is theoretical; they graphically illustrate the inferential structure of the grouping theory, showing exactly what the competing hypotheses are, and how one is chosen over all the others by the perceiver.

An Efficient Procedure

This section presents an efficient procedure for computing interpretations without constructing the full interpretation space. The motivation of the algorithm derives directly from the logical structure of the interpretation space.

Above, the partial order on interpretations was treated as an abstract, formal entity, expressing a preference ranking among mathematical objects. Alternatively, the very same partial order can be interpreted *procedurally* as the temporal sequence in which interpretations should be evaluated. This interpretation is in fact made possible by the logical structure itself, and in particular by the Genericity Constraint.

For a given interpretation to be preferred, the observed configuration must be generic in it, which means that all interpretations beneath it (deeper) in the partial order must not hold. Hence it makes sense to evaluate more preferred interpretations first, and then stop when one succeeds. When an interpretation fails, its upper neighbors are evaluated next (in no particular order). In this manner, when an interpretation does succeed, it can be assumed to be generic, because it will be known

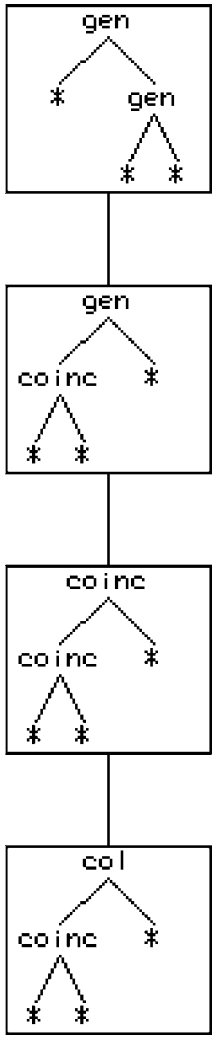


Figure 5. The interpretation space for $n = 3$ dots.

that all more preferred interpretation will already have failed, thus satisfying the definition of genericity. Note that only one interpretation on a row of a particular lattice can succeed. If two interpretations T_1 and T_2 hold, then so does their meet (greatest lower bound) $T_1 \wedge T_2$, which is a node on a lower row, and which thus by definition will have already failed.

In this way, all interpretations in the partial order can be evaluated in order from the bottom of the partial order to the top, and whichever succeeds first is guaranteed to be preferred. In practice, interpretations need not be evaluated completely; a much simpler technique is to recurse dot by dot, as follows.

The base case is two dots, which are assigned an interpretation by evaluating interpretations in reverse

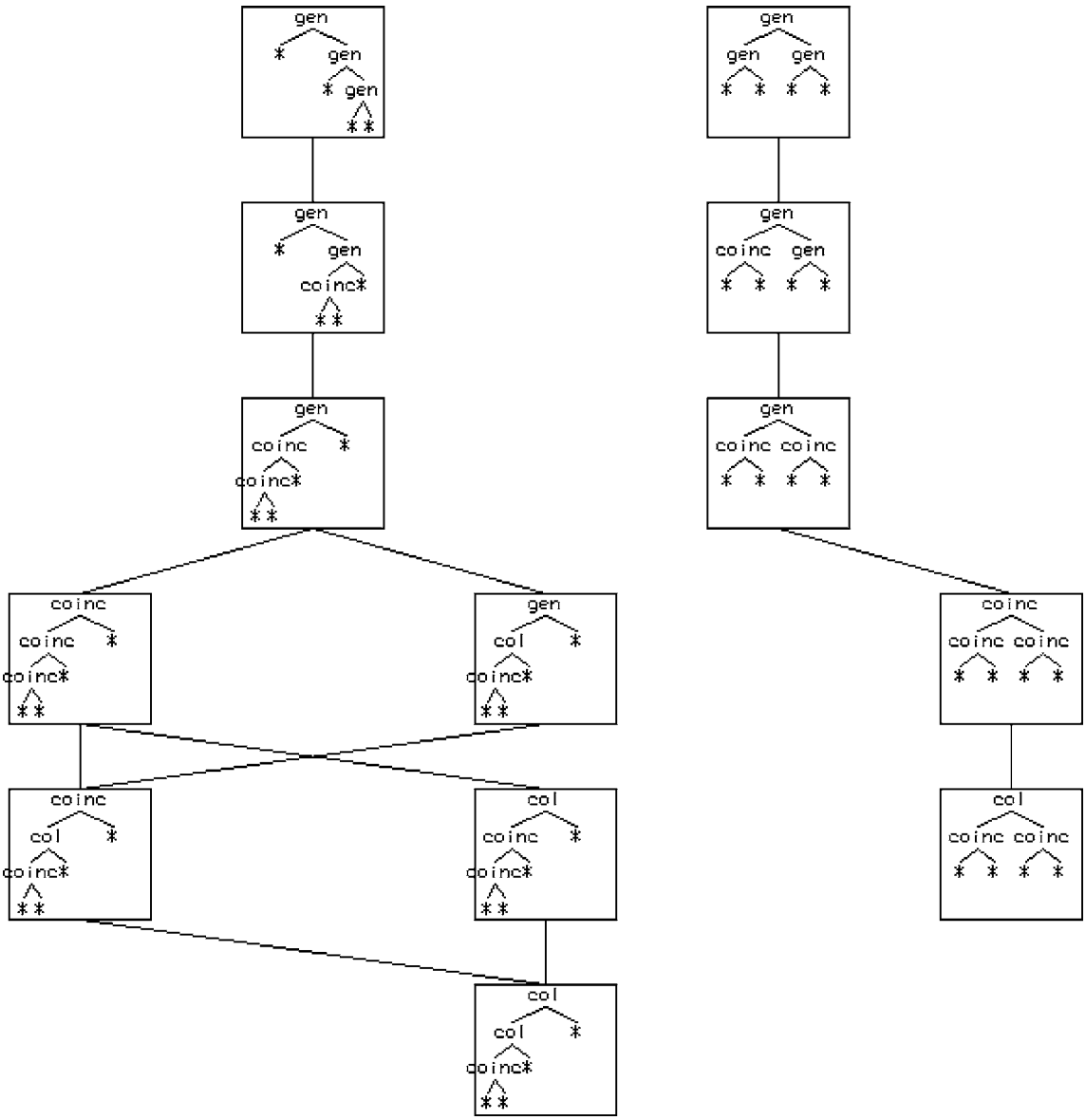


Figure 6. The interpretation space for $n = 4$ dots.

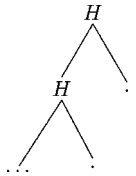
order of depth. Using the dot regularity set defined above as an example, if two dots are `coincident`, they are assigned the tree (2). Otherwise, the two dots are assigned the generic tree (3). The recurse case of $n + 1$ dots $\{x_1, \ldots, x_{n+1}\}$ is handled as follows:

(Algorithm)

Step 1. Find the minimal interpretation T for $x_1 \ldots x_n$.

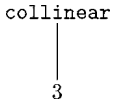
Step 2. Join x_{n+1} to T in as preferred a manner as possible.

Step 1 is recursive. Step 2 is accomplished as follows. A *chain* is a tree such as



(9)

in which a head term H has itself as a child. The length of the chain is the number of dots in its scope. Chains appear frequently in high-depth solutions, so it is convenient to adopt an abbreviated notation, in which e.g.



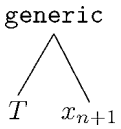
(10)

is a chain of 3 collinear dots. Like all parse trees, each chain has a depth. Now, in joining x_{n+1} to the preexisting tree T, it makes sense to focus on high-dimension subtrees, and in particular, chains. In keeping with procedural interpretation of the partial order, these subtrees should be considered in reverse order of depth, from deepest to least deep. So the joining step (2) is carried out as follows:

(Joining step [Step 2])

1. Choose the chain $C \subset T$ with maximum depth;
2. Join x_{n+1} to C in the maximum depth way.

This last step simply follows the main model lattice of Fig. 3. Finally if no regularities hold between x_{n+1} and any chain C of T, then the configuration is assigned the tree



(11)

Thus the procedural interpretation of the partial order is observed at two levels: preexisting models are considered in order of preference, and models of their relationship to the new dot are considered in order of preference. The result is a global interpretation that is maximally preferred at each stage of its construction. It is not, however, guaranteed to be the globally maximum-depth solution. For one thing, the solution found depends on the sequence in which the dots are considered. Nevertheless, in practice on moderately structured configurations such as those arising from natural images, and simply considering image elements in left-to-right order, it typically finds the global optimum or some slight variant of it (see examples below).

Complexity

Because the procedure terminates with success but continues on failure (which corresponds to lack of structure in the configuration) processing time is highly dependent on the degree of structure in the configuration. The worst case is n mutually generic dots. In this case the program attempts to join the $n+1$-th dot with each of the existing n dots, for a total of $1+2+\cdots+n = O(n^2)$ processing steps. At the other extreme, a completely ordered configuration such as a collinear chain will take only n steps, as each new dot is added in one step to the single existing subtree. Hence the procedure is efficient ($O(n^2)$) in general, and extremely rapid ($O(n)$) when the image is highly structured. In practice the procedure takes only a few seconds on a Sparc Station 2 even on randomly generated displays of hundreds of dots, and is effectively instantaneous on smaller, highly structured configurations such as Fig. 8.

Identification of Salient Structure

The fact that there is an absolute measure of the degree of "goodness" for each interpretation (namely depth) presents several distinct advantages over conventional techniques. One is that highly regular structures embedded within the image are identified as a side-effect. Such structures appear within the overall interpretation as high-depth subtrees and chains. This means that even though the overall solution is a global description of the entire configuration, it automatically articulates internal structures that ought to be regarded as coherent, providing figure/ground separation and effective discovery of object-like components of the scene—i.e., "salient" structure (cf. Shashua and Ullman, 1988; Williams and Thornber, 1999). Again it is to be emphasized that MM theory was not explicitly designed to solve this problem, but rather the more general problem of grouping and perceptual organization; yet the solution, and its complexity, compare reasonably favorably with techniques specially designed for this problem. Examples are shown below.

Examples and Experiments

Figure 7 shows the minimal interpretation of the configuration from Fig. 2. In this simple example, the interpretation closely corresponds to the intuitive percept of a dot chain and a dot cluster. Figure 8(a) shows a slightly more difficult example, this time constructed from edge elements instead of dots. (Edge elements are items with

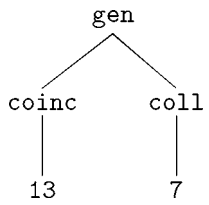


Figure 7. The qualitative parse (maximum-depth interpretation) of the dot configuration from Fig. 2. The tree correctly finds a `coincident` cluster of 13 dots and a `collinear` chain of 7 dots.

a location and orientation but no spatial extent.) Intuitively, one sees two intersecting chains, and the minimal interpretation indeed gives a `generic` head term over two `collinear` chains. The tree is given in two forms, once explicitly (Fig. 8(b)) and once (Fig. 8(c)) in schematic form (convenient for large trees). In the schematic depiction of the tree, elements that fall in the scope of a common chain are connected by line segments.

Figures 9 and 10 show more difficult examples, drawn from natural images. In these figures, the input to the algorithm is a compound image created by combining an image of a real object (panel a) with an image of a real background texture (panel b). Panel (c) shows the combined edge image. The advantage of using a compound image for evaluation purposes is that it allows an objective measure of whether the algorithm parsed the image "correctly;" because we know which edge elements really derived from the object and which from the background, we can determine whether the algorithm correctly segregated the two sources (Williams and Thornber, 1999). This allows

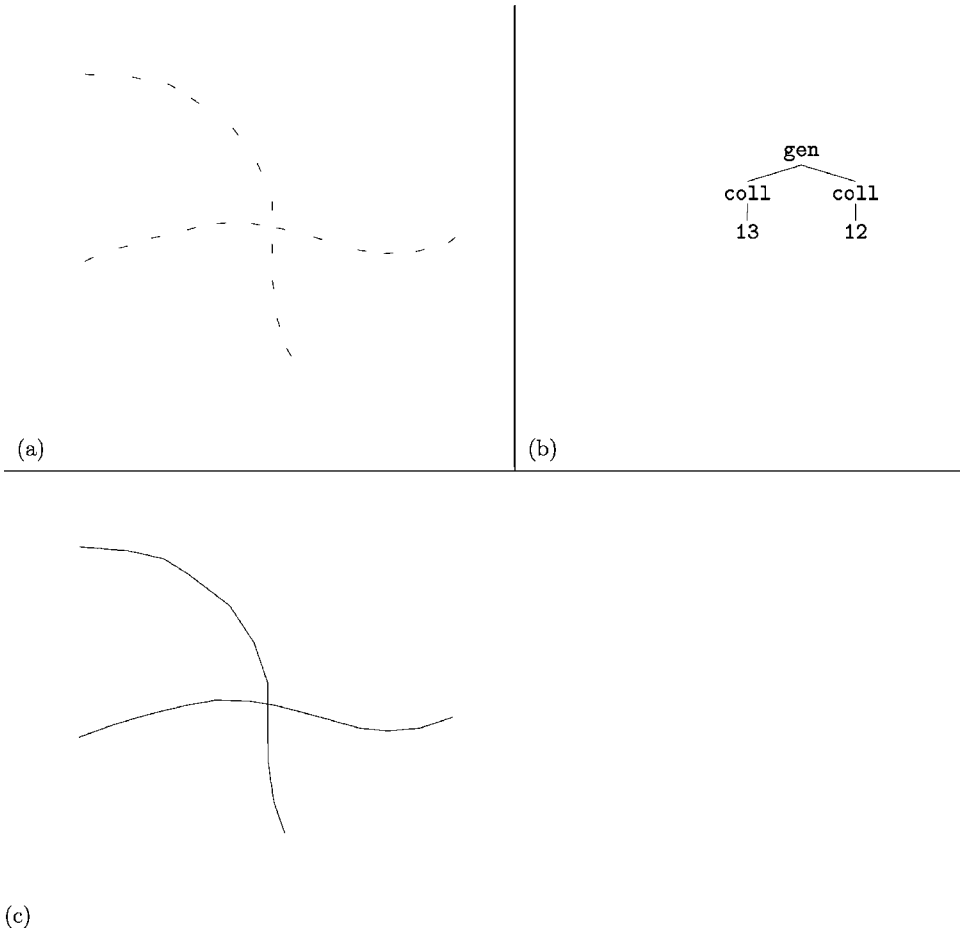


Figure 8. (a) A configuration of edge elements. (b) The maximum-depth parse tree, consisting of two `collinear` chains. (c) The same tree shown schematically (explained in text).

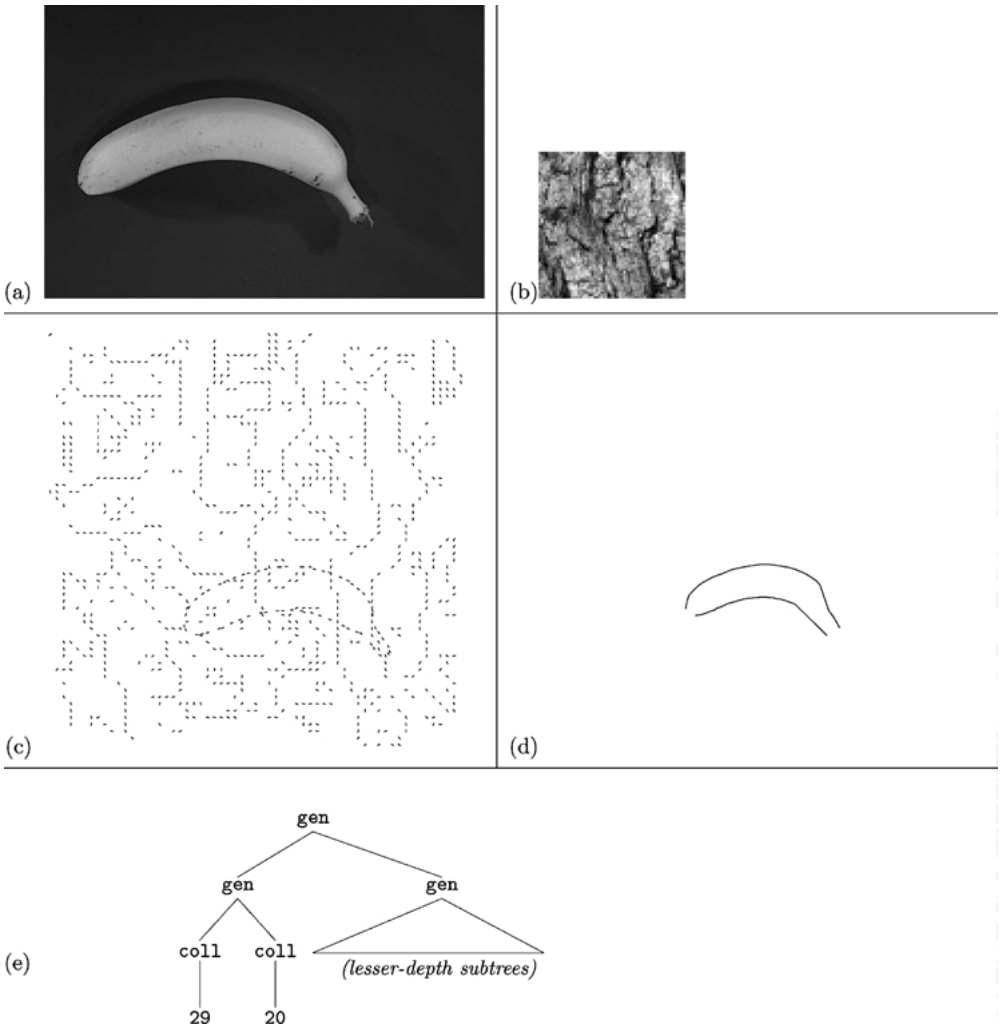


Figure 9. (a) Image of a banana. (b) Image of bark background. (c) Combined edge image of banana embedded in bark background. The banana and the background contribute 61 and 842 edge elements respectively. (d) The two maximum-depth subtrees in the maximum-depth interpretation of the combined image (shown schematically). (e) The maximum-depth parse tree (abbreviated).

more objective evaluation of the method that is possible with a complete natural images, where performance must be evaluated subjectively. The rationale for using a natural texture rather than random noise elements is that natural textures contain collinearities and other regular structure, thus presenting a more difficult and realistic foil for the algorithm.

Figure 9 combines an image of a banana[8](Fig. 9(a)) and an image of a bark texture (Fig. 9(b)), extracting edges using a standard operator, and combining the two resulting edge maps (Fig. 9(c)). In the combined edge map, the banana can still be discerned by eye, but the discrimination is not easy. Figure 9(d) and 9(e) show the minimal interpretation of the combined edge map. Figure 9(d) shows the two[9]maximum-depth subtrees (schematically, as explained above), while Fig. 9(e) shows (part of) the maximum-depth parse tree explicitly. The minimal interpretation picks out the banana fairly effectively. Figure 10 shows a similar example using a peach with a terrain background. In both the banana/bark and peach/terrain cases, the minimal interpretation plainly corresponds to the intuitive interpretation of "object plus background," with the tree explicitly (although imperfectly) indicating which edge elements belong to the "object."

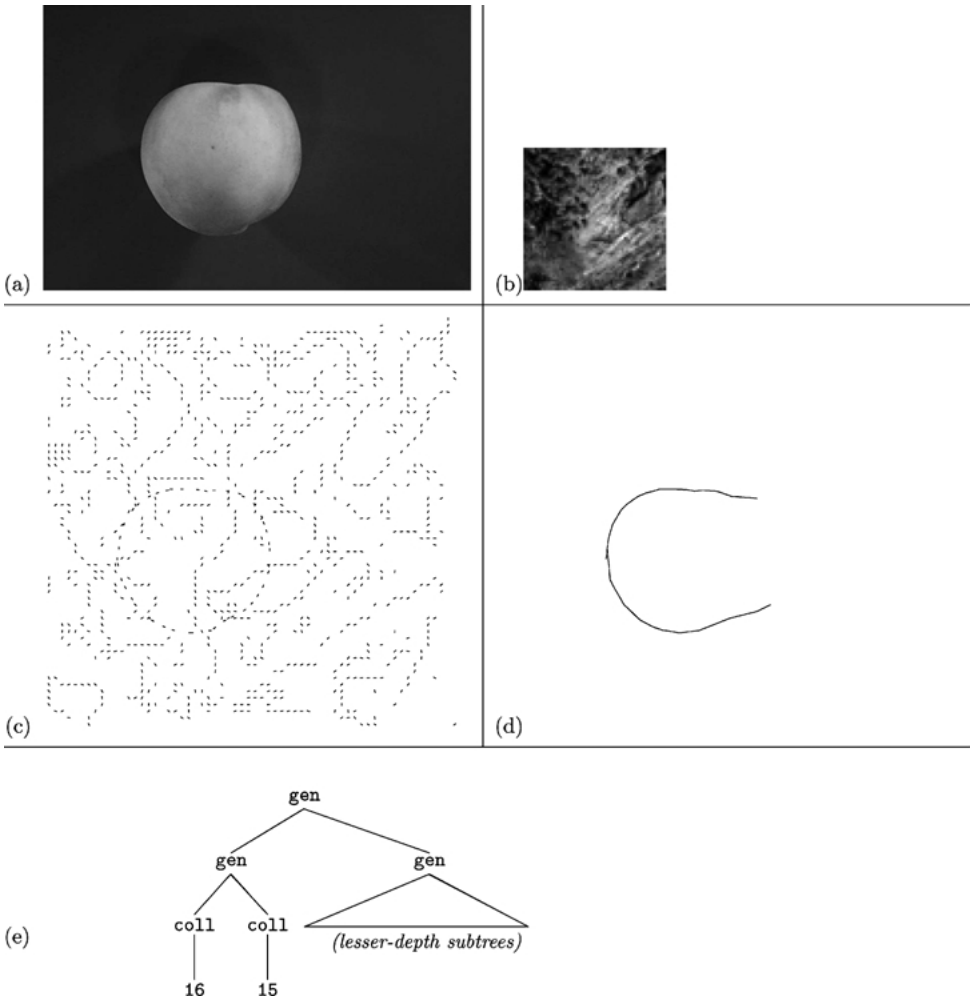


Figure 10. (a) Image of a peach. (b) Image of terrain background. (c) Combined edge image of peach embedded in terrain background. The peach and the background contribute 44 and 840 edge elements respectively. (d) The two maximum-depth subtrees in the maximum-depth interpretation of the combined image (shown schematically). (e) The maximum-depth parse tree (abbreviated).

Parametric Experiments

The fact that compound images allow objective performance evaluation makes it possible to conduct parametric experiments on the algorithm. In the following experiments, the same compound image (peach + terrain or banana + bark) is presented to the algorithm at a variety of values of the two main parameters, the proximity threshold $\theta_{\text{proximate}}$ and the collinearity threshold $\theta_{\text{collinear}}$. The upper row of Fig. 11 (peach + terrain) and Fig. 12 (banana + bark) then gives the proportion of edge elements correctly classified as a function of the parameter value over a large range of parameter settings.

One obstacle to objectively evaluating performance in an object-extraction paradigm is that extracting greater numbers of elements will always tend to catch more true object edges, but at the expense of also increasing "false positives," i.e. elements identified as object but actually background (cf. Williams and Thornber, 1999). This creates a problem here because the number N of highest-depth subchains promoted as "object" is arbitrary. The performance measure d' from Signal Detection theory (Green and Swets, 1966) solves this problem, giving a measure of true sensitivity that is independent of the threshold for responding "yes" (i.e., here independent of how many subchains are promoted). This measure is more revealing than

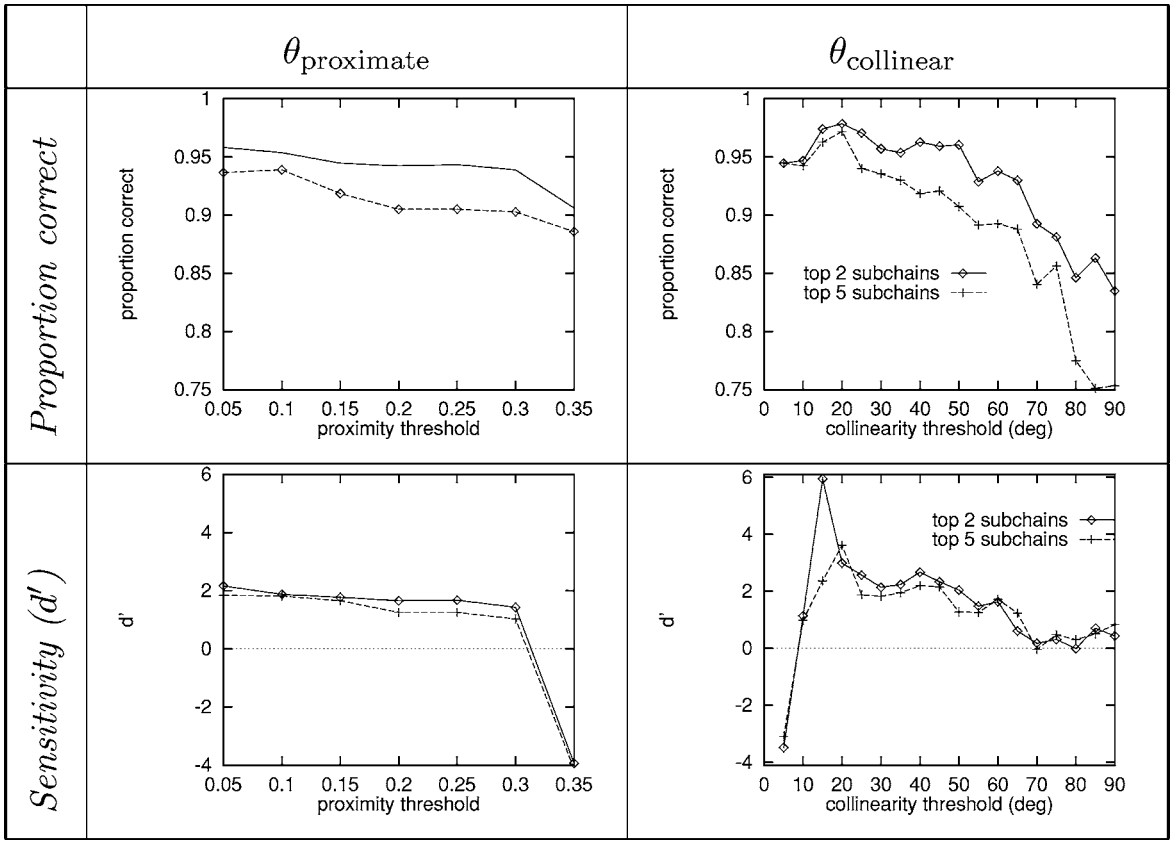


Figure 11. Results for peach + terrain images, showing performance as a function of the choice of parameters. The upper row gives proportion correct classification of edges into 'object' (peach) vs. 'background' (terrain), as a function of proximity threshold $\theta_{\text{proximate}}$ (left column) or collinearity threshold $\theta_{\text{collinear}}$ (right column). Here and elsewhere values for $\theta_{\text{proximate}}$ are given as a ratio of total image size. The lower row gives true sensitivity (d'), again a function of $\theta_{\text{proximate}}$ or $\theta_{\text{collinear}}$.

simply counting false positives in measuring the true separate of signal from noise. The application of such techniques to computer vision has been pioneered in Bowyer and Phillips (1998). The lower row of Figs. 11 and 12 gives sensitivity d' as a function of parameter settings.

The results (Figs. 11 and 12), on both image sets and using both performance measures, show that (a) the algorithm's performance is objectively good, with proportion correct about 0.95 and d' about 2 at the best parameter settings; and (b) this good performance is robust over a wide range of parameter settings, only breaking down at extreme values (e.g. collinearity threshold greater than 60°). Moreover these experiments serve to identify optimal parameter settings in a manner that is independent of subjective judgments about the quality of the results.

Williams and Thornber (1999) mounted an extensive comparison of salience-extraction methods, in which the various methods were compared using a common performance measure, proportion false positives (that is, the percent of edges intepreted as object that were in fact background). Figure 13 shows a similar analysis of the current algorithm, averaging over performance in the peach + terrain and banana + bark compound images. The algorithm's performance is shown at the same levels of signal-to-noise ratio used in their study, for ease of comparison with their analysis. As in their study, in each case the noise image has been randomly undersampled to achieve the desired signal-to-noise ratio, and the object has been scaled to half the size of the noise image. Results are plotted for three different levels of N (the number of sub-trees promoted as "object"). The algorithm's

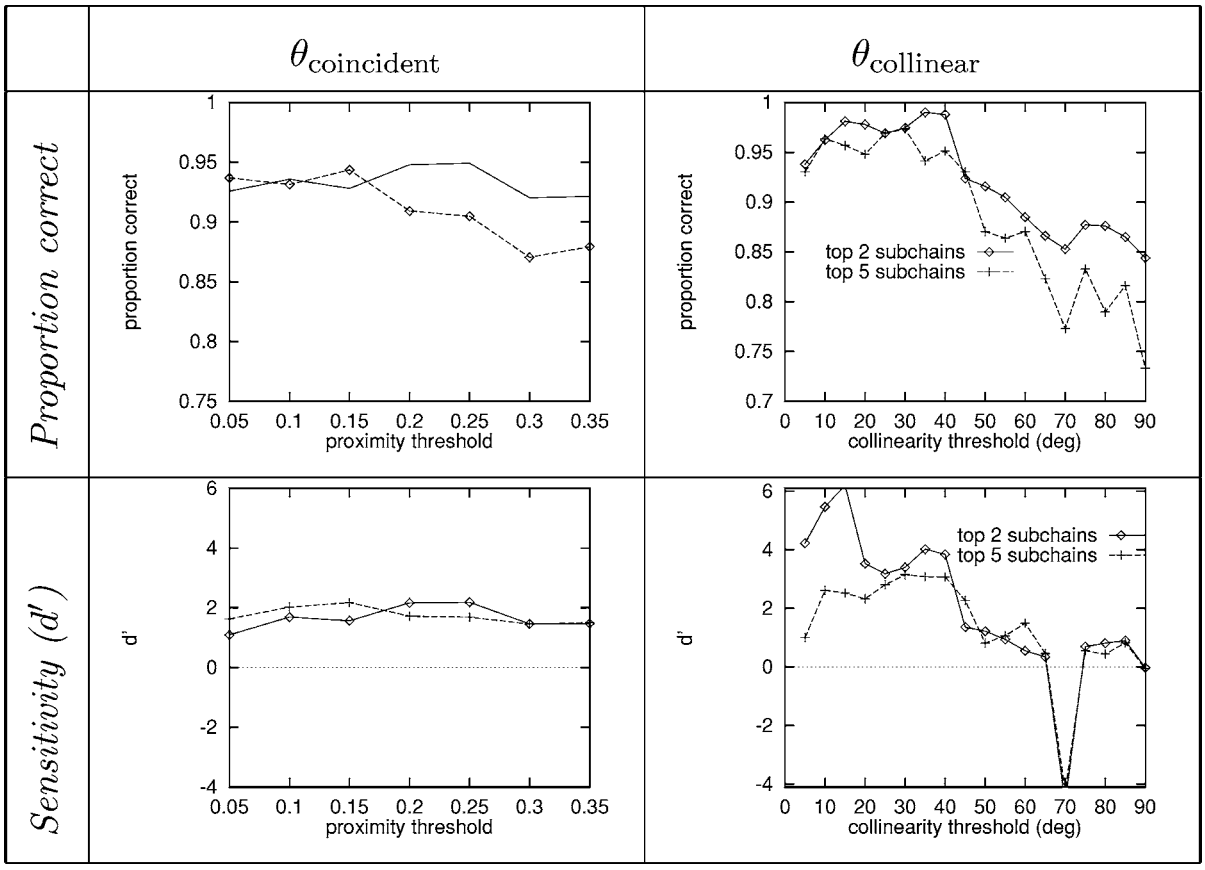


Figure 12. Results for banana + bark images, showing performance as a function of the choice of parameters (see previous caption).

performance is slightly worse than that of Williams and Thornber (1999)'s own algorithm (the winner in their analysis) but better than all the others considered, including the algorithms of Shashua and Ullman (1988), Hérault and Horaud (1993), Sarkar and Boyer (1996), Guy and Medioni (1996), and Williams and Jacobs (1997). However these comparisons should be interpreted with caution, as the current algorithm has not been evaluated in precisely the same way nor on precisely the same images (actually, a subset) as in Williams and Thornber (1999)'s scrupulously neutral comparison.

Finally, Figs. 14 and 15 show results on complete natural images using optimal parameter settings selected based on the results of the experiments. Figure 14 shows simple silhouette images of tools including occlusions. The contours produced by the algorithm all derive from either one tool or the other. Figure 15 shows a far more complex natural image of peppers, with many objects and many occlusions. Here performance

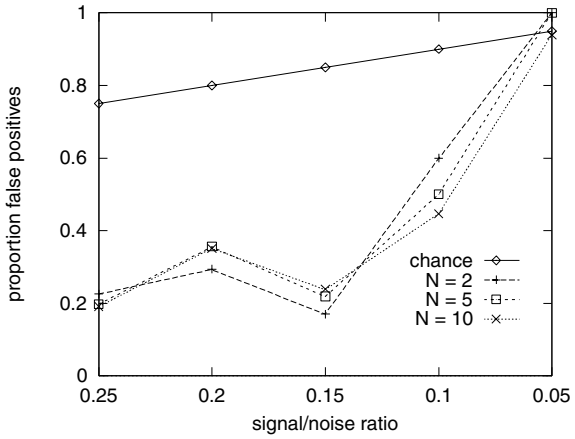


Figure 13. Performance of the algorithm (averaging over peach + terrain and banana + bark images) as a function of signal-to-noise ratio. The graph shows proportion false positives at several levels of N (the number of subtrees promoted as object), and also shows chance performance. This level of performance compares favorably with most, but not all, of the algorithms discussed in Williams and Thornber (1999) (see text for discussion).

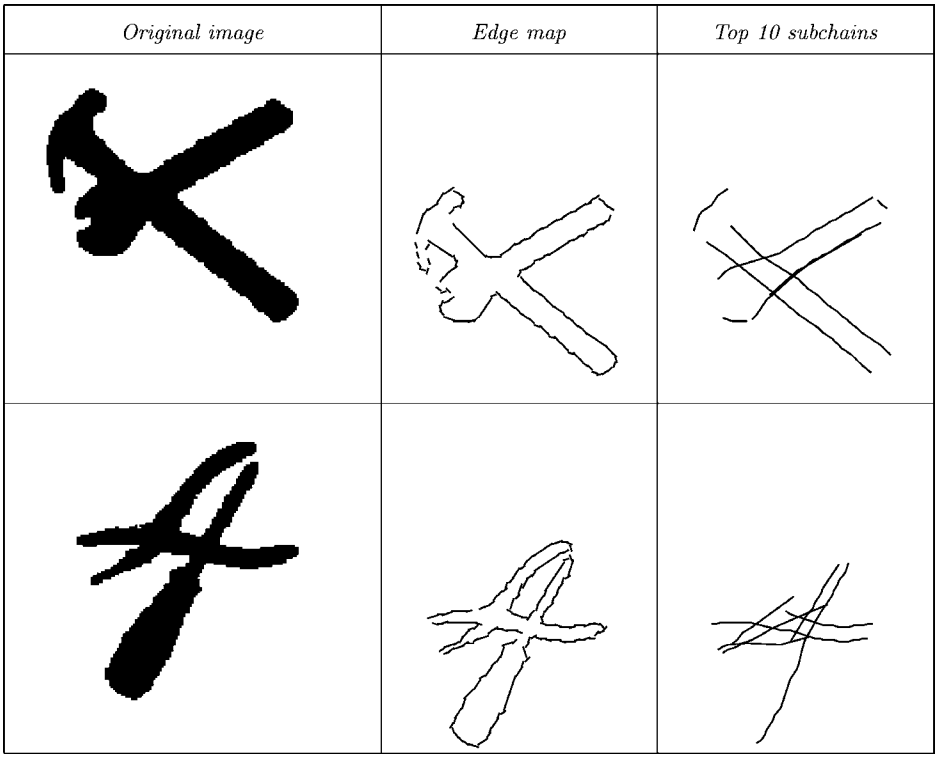


Figure 14. Results for two silhouette images with occlusion (using $\theta_{\text{proximate}} = 0.10, \theta_{\text{collinear}} = 25^\circ$) showing 10 maximal-depth subchains (right column). In both cases the two tools are correctly separated in the minimal model, with each maximum-depth chain originating from only one tool.

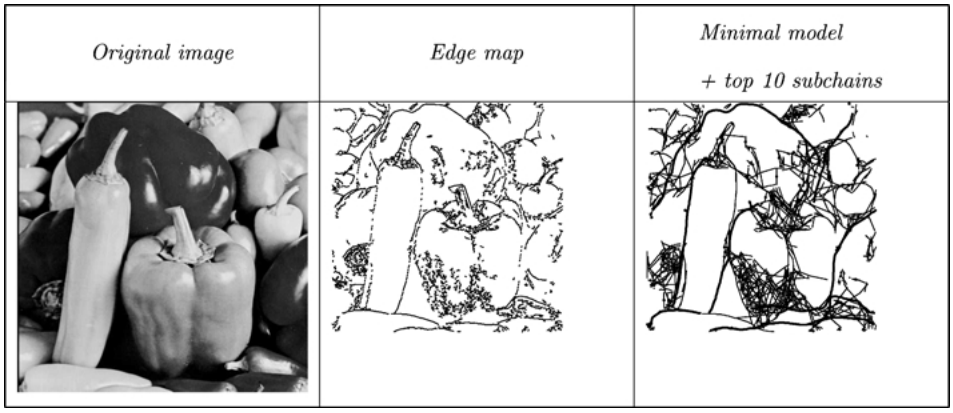


Figure 15. Results for a cluttered natural image of peppers (using $\theta_{\text{proximate}} = 0.15, \theta_{\text{collinear}} = 25^\circ$), showing (right column) the complete minimal model along with the 10 maximum-depth subhchains emphasized.

could certainly be improved if more regularities were added to the premise set (e.g., region-based ones); the results displayed derive only from the contour-based regularities `coincident` and `collinear`.

Discussion

The closest comparison with Minimal Model theory in the computational vision literature is the use of Normalized Cuts by Shi and Malik (2000; see also the related work of Amir and Lindenbaum, 1998), which also creates a hierarchical representation of the relations among image elements. In Shi and Malik's technique, the adjacency graph derived from element affinities is decomposed in an optimal way (i.e., one that minimizes the edge weight between subgraphs while maximizing the weight within each subgraph, thus yielding to heterogeneous but internally homogeneous image regions). Then the resulting subgraphs are themselves recursively decomposed, and so forth, resulting in a tree representing the hierarchy of relations among regions in the original graph. Like the parse tree of MM theory, the relations in this tree are hierarchical. But unlike MM theory, even at higher levels, the relations depend on the affinities among image elements at the lowest level (because the decomposition results from cuts in the original adjacency graph created by those affinities). This means that the only Gestalt relations or predicates in effect are those that operate at the level of individual image items, such as the commonly-cited principles of proximity and good continuation. This precludes any predicates that operate at the level of a whole aggregation of elements, such as symmetry, closure, and the many other non-local principles identified by the Gestaltists (and confirmed by modern data; see discussion above). Broadly speaking, then, the representation generated in Minimal Model theory can be regarded as similar to that in Shi & Malik's theory, but augmented by the possibility of these other more abstract predicates and Gestalt-like principles. Unfortunately, this means that an elegant and efficient top-down computation of the optimal decomposition, such as the generalized eigenvalue computation of the segmented graph proposed by Shi & Malik, is no longer possible. In MM theory, in order to provide arguments to these more abstract predicates, lower-level aggregations must be computed first. This leads to the admittedly less elegant greedy computation of the Minimal Model described above. As discussed, this algorithm is intended merely to show that a reasonable approximation of the Minimal Model can be computed reasonably efficiently.

Another relevant comparison for the current work is the cluster of recent approaches in which Bayesian, statistical, or voting methods are employed to give quantitative estimates of "salience" or goodness of curves connecting visual items (e.g. Cox et al., 1993; Guy and Medioni, 1996; Medioni et al., 2000; Williams and Jacobs, 1997). In these approaches, quantitative details of inter-element relations (e.g., where the observed angles fall in a distribution of angles defined by some stochastic model) are combined to lead to quantitative estimates of the implied curve. These approaches generally seem to represent a very effective way of estimating the goodness of curves, and indeed as discussed above bear a close similarity to the Bayesian machinery of contour integration that human observers seem to employ (Feldman, 2001b). In the simple case of a chain of dots, the relatively simple additive cue combination of thresholded affinities employed in the current research is, in fact, simply a coarse approximation to the more fine-tuned estimates provided by these other approaches—which is presumably why the results are generally similar. But as discussed above, in Minimal Model theory this salient-curve extraction is simply a side-effect of a larger organizational structure designed to provide a high-level, hierarchical, qualitative structural description of the entire configuration, and thus represents different goals and motivations from these other approaches. Again, these different goals reflect the theory's origins as an attempt to capture human perceptual organization computationally.

Summary and Conclusions

Minimal Model theory is an application of logical machinery to the grouping problem. Interpretations, defined as logical objects, are ranked (partially-ordered) by their degree of regularity or accidentalness. Among all interpretations that a given configuration satisfies, the most preferred is the minimum in this partial order, i.e. the one with maximum depth. This interpretation accounts as completely as possible for what otherwise would be highly suspicious coincidences in the image configuration. In this sense the interpretation amounts to the best possible "explanation" of the processes that gave rise to the image elements—in a very literal sense, the "simplest" (algebraically

minimal) interpretation of the image configuration. The algorithm described above establishes that logically optimal interpretations can be computed efficiently.

Computational grouping is widely perceived as a field in need of fundamentally new ideas. The current approach offers has a number of benefits:

- The interpretation is perceptually compelling. Figures 9 and 10 in particular make the case that the minimal interpretation is intuitively correct. As the human visual system is still the gold standard, this is clearly a central goal of computational research. Moreover, as the experiments suggest, the minimal interpretation is objectively correct with high probability.
- The interpretation is global in nature, in that it represents an optimal model of the entire scene. In effect, the goal of MM theory is to formulate a logic whereby the many local interpretation preferences throughout an image may be optimally combined into a single global percept, a problem sometimes called *cooperativity* in the psychological literature (Julesz, 1995; Kubovy and Wagemans, 1995), and generally considered unsolved.
- The interpretation is highly robust to small changes in the configuration. In general, image changes that do not lead to any change in the truth value of a regularity predicate do not change the minimal model at all (by definition). Image changes that do change regularity valuations will change the precise structure of trees consistent with the image, but generally will *not* affect which tree wins (i.e. is minimal), except in very close cases, which are rare with natural images. More precisely, random portions of the image will generally be described in the minimal tree by subtrees with random structure and usually low depth (because each degree of depth corresponding to one random accidental regularity). When such portions change to a new random pattern the parse tree will likewise change in a random fashion. However *non*-random portions of the image, which obey regularities generically—such as "objects"—will continue to be described by high-depth subtrees in a relatively stable manner (see Feldman, 1999). Hence the overall structure of the winning interpretation—e.g., which parts of the image are grouped together, and which part is regarded as the "figure"—survives both the addition of background noise and other varieties of image degradation. This is not a "magic trick," or absurdly strong claim; it is a direct result of the *qualitativeness* of the image description. By aiming only for a qualitative, categorical description, one based solely around the recognition of stable structures, the theory gives up on the aim for precise descriptions. The representation inherently gains in stability what it gives up in quantitative precision.
- The interpretation is not just a processed image, but rather gives qualitative symbolic descriptions to distinct parts of the scene. Many grouping techniques simply yield an image in grouped image regions are in some way enhanced. Such a representation is still an image, and needs some kind of additional processing before any kind of qualitative or symbolic description can be achieved. Parse trees constitute an initial form for such description. As suggested by the examples above, the structure of the trees bears a close analogy to the type of verbal description a human observer would intuitively apply to the scene.
- The framework is not specific to dot grouping, and (as demonstrated above) is easily extended to other primitives (edges, line segments, etc.) or other regularity sets (e.g. adding smoothness, cocircularity, and more abstract predicates, etc.; although admittedly this type of extension remains untested). This makes it possible in principle to extend the approach to different problems in perceptual organization while preserving the essential logic of the approach, instead of inventing new ad hoc methods for each new problem. This mirrors the view often expressed in the psychological literature that human perceptual organization derives from a small number of uniform core principles, e.g. the Gestalt notion of *Prägnanz*.

Minimal Model theory was conceived not as an alternative to conventional computational grouping methods, but rather as an attempt to give a formal, rigorous account of human grouping intuitions. In a sense, as discussed above, it attempts to solve a different (perhaps complementary) problem from that attacked by other grouping work. Because the theory turns out to have an efficient computational implementation, as demonstrated above, it has the promise of pointing the way to algorithms that can effectively mimic the flexibility and power of the human visual system.

Acknowledgments

I am grateful to Allan Jepson, Yakov Keselman, Alan Mackworth, Ray Reiter, Whitman Richards, Lance Williams, and three anonymous reviewers for helpful discussions and comments; to Lance Williams for providing the images used in Figs. 9 and 10; and to Ali Shokoufandeh for assistance in preparing the images used in Figs. 14 and 15. Portions of this paper appeared as "Efficient regularity-based grouping," *CVPR*, 1997. Supported in part by NSF SBR-9875175.

Notes

1. In geometry, the codimension is the difference in dimension between an object and the space in which it is embedded (see Poston and Stewart, 1978). Here, it is the difference in dimension between a regular model and the generic case. For purposes of this paper, depth and codimension may be regarded as synonymous. Feldman (1997b) discusses the distinction and gives sufficient conditions for regarding them as equivalent.
2. The predicate `coincident` might be more accurately rendered in English as "proximate," since it simply specifies that two dots fall nearby each other—just as other predicates (e.g. `collinear`) also have a resolution or tolerance parameter. The name `coincident`, like the name `collinear`, is preferable only in that it emphasizes the dimensional "accident" (i.e., the coincidence of spatial dimensions) in the perfect case, which underlies the predicate's status as a regularity. In any case, the name of the predicate is arbitrary and serves only as a mnemonic.
3. The dependence here on negation requires the application of the negation-as-failure rule, which as readers familiar with Logic Programming will know, can create difficulties in the semantics. Tightening up the semantics requires the Closed World Assumption (Reiter, 1978); again see Feldman (1997b).
4. Here we ignore the degenerate case of two perfectly coincident dots.
5. A distributive lattice is an algebraic structure defined by two operations $\wedge$ (*meet*) and $\vee$ (*join*) obeying the distributive equalities $a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$ and $a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$ for all a, b, c. The partial order $\leq$ under discussion here is an algebra under the mapping $T_1 \leq T_2$ iff $T_1 \wedge T_2 = T_1$ and $T_1 \vee T_2 = T_2$. See Davey and Priestley (1990) for an introduction to lattice theory and distributivity, and see Feldman (1997b) for a proof and discussion of the distributivity of grouping interpretation spaces.
6. See Williams and Thornber (1999) for a discussion of the importance of contour orientation in defining collinearity.
7. See note 5 above.
8. I am grateful to Lance Williams for providing these images.
9. Because of the way the algorithm works (i.e., considering elements from left to right, and building the tree in a greedy fashion), chains tend not to be able to round leftmost and rightmost corners well. Hence a closed curve usually shows up as two chains rather than one. The two chains can easily be combined in a trivial post-processing step to yield a single closed curve. The "pure" solution is shown here for clarity of exposition.

References

Amir, A. and Lindenbaum, M. 1998. A generic grouping algorithm and its quantitative analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:168–185.

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

Baylis, G. and Driver, J. 1993. Visual attention and objects: Evidence for hierarchical coding of location. *Journal of Experimental Psychology: Human Perception and Performance*, 19(3):451–470.

Bennett, B.M., Hoffman, D.D., and Prakash, C. 1989. *Observer Mechanics: A Formal Theory of Perception*. London: Academic Press.

Binford, T. 1981. Inferring surfaces from images. *Artificial Intelligence*, 17:205–244.

Boselie, F. and Wouterlood, D. 1989. The minimum principle and visual pattern completion. *Psychological Research*, 51:93–101.

Bowyer, K.W. and Phillips, J. (eds.). 1998. *Empirical Evaluation Techniques in Computer Vision*. IEEE Computer Society Press.

Compton, B.J. and Logan, G.D. 1993. Evaluating a computational model of pereptual grouping by proximity. *Perception & Psychophysics*, 53(4):403–421.

Cox, I.J., Rehg, J.M., and Hingorani, S. 1993. A bayesian multiple hypothesis approach to contour segmentation. *International Journal of Computer Vision*, 11:5–24.

Davey, B. and Priestley, H. 1990. *Introduction to Lattices and Order*. Cambridge: Cambridge University Press.

Elder, J. and Zucker, S. 1993. The effect of contour closure on the rapid discrimination of two-dimensional shapes. *Vision Research*, 33(7):981–991.

Feldman, J. 1996. Regularity vs. genericity in the perception of collinearity. *Perception*, 25:335–342.

Feldman, J. 1997a. Curvilinearity, covariance, and regularity in perceptual groups. *Vision Research*, 37(20):2835–2848.

Feldman, J. 1997b. Regularity-based perceptual grouping. *Computational Intelligence*, 13(4):582–623.

Feldman, J. 1997c. The structure of perceptual categories. *Journal of Mathematical Psychology*, 41:145–170.

Feldman, J. 1999. The role of objects in perceptual grouping. *Acta Psychologica*, 102:137–163.

Feldman, J. 2001a. *Bayes and the Simplicity Principle in Perception*. Manuscript under review.

Feldman, J. 2001b. Bayesian contour integration. *Perception & Psychophysics*, 63(7):1171–1182.

Freeman, W.T. 1994. The generic viewpoint assumption in a framework for visual perception. *Nature*, 368:542–545.

Gilchrist, A.L. and Jacobsen, A. 1989. Qualitative relationships are decisive. *Perception and Psychophysics*, 45(1):92–94.

Glass, L. 1969. Moiré effects from random dots. *Nature*, 223:578–580.

Green, D.M. and Swets, J.A. 1966. *Signal Detection Theory and Psychophysics*. New York: Wiley.

Guy, G. and Medioni, G. 1996. Inferring global perceptual contours from local features. *International Journal of Computer Vision*, 20(1/2):113–133.

Iverson, L. and Zucker, S. 1995. Logical/linear operators for image curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(10):982–996.

Jacobs, D. 1996. Robust and efficient detection of salient convex groups. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:23–37.

Jepson, A. and Mann, R. 1999. Qualitative probabilities for image interpretation. In *Proceedings of the International Conference on Computer Vision*, vol. II, pp. 1123–1130.

Jepson, A. and Richards, W.A. 1991. *What is a Percept*? Occasional Paper No. 43. MIT Center for Cognitive Science.

Jepson, A. and Richards, W.A. 1992. What makes a good feature? In *Spatial Vision in Humans and Robots*, L. Harris and M. Jenkin (Eds.), Cambridge University Press, pp. 89–125.

Kanizsa, G. 1979. *Organization in Vision: Essays on Gestalt Perception*. New York: Praeger Publishers.

Kelly, A.R. and Hancock, E.R. 2000. Grouping line-segments using eigenclustering. In *Proceedings of the 11th British Machine Vision Conference*, pp. 586–295.

Kitazaki, M. and Shimojo, S. 1996. 'Generic view principle' for three-dimensional-motion perception: Optics and inverse optics of a moving straight bar. *Perception*, 25:797–814.

Kovacs, I. and Julesz, B. 1993. A closed curve is much more than an incomplete one: Effect of closure in figure-ground segmentation. *Proceedings of the National Academy of Science*, 90:7495–7497.

Kubovy, M. 1994. The perceptual organization of dot lattices. *Psychonomic Bulletin and Review*, 1(2):182–190.

Kubovy, M., Holcombe, A.O., and Wagemans, J. 1998. On the lawfulness of grouping by proximity. *Cognitive Psychology*, 35:71–98.

Kubovy, M. and Wagemans, J. 1995. Grouping by proximity and multistability in dot lattices: A quantitative gestalt theory. *Psychological Science*, 6(4):225–234.

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

Markman, A.B. and Gentner, D. 1993. Structural alignment during similarity comparisons. *Cognitive Psychology*, 25:431–467.

Medin, D.L., Goldstone, R.L., and Gentner, D. 1990. Similarity involving attributes and relations: Judgments of similarity and difference are not inverses. *Psychological Science*, 1(1):64–69.

Medioni, G., Lee, M.-S., and Tang, C.-K. 2000. *A Computational Framework for Segmentation and Grouping*. Amsterdam: Elsevier.

Nakayama, K. and Shimojo, S. 1992. Experiencing and perceiving visual surfaces. *Science*, 257:1357–1363.

Palmer, S.E. 1977. Hierarchical structure in perceptual representation. *Cognitive Psychology*, 9:441–474.

Palmer, S.E. 1978. Structural aspects of visual similarity. *Memory & Cognition*, 6(2):91–97.

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

Palmer, S.E. 1992. Common region: A new principle of perceptual grouping. *Cognitive Psychology*, 24:436–447.

Parent, P. and Zucker, S.W. 1989. Trace inference, curvature consistency, and curve detection. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 11(8):823–839.

Pizlo, Z., Salach-Golyska, M., and Rosenfeld, A. 1997. Curve detection in a noisy image. *Vision Research*, 37(9):1217–1241.

Pomerantz, J.R. 1986. Visual form perception: An overview. In *Pattern Recognition by Humans and Machines, vol. 2: Visual Perception*. Orlando, FL: Academic Press.

Pomerantz, J.R. and Pristach, E.A. 1989. Emergent features, attention, and perceptual glue in visual form perception. *Journal of Experimental Psychology: Human Perception and Performance*, 15(4):635–649.

Pomerantz, J.R., Sager, L.C., and Stoever, R.J. 1977. Perception of wholes and their component parts: Some configural superiority effects. *Journal of Experimental Psychology: Human Perception and Performance*, 3(3):422–435.

Poston, T. and Stewart, I.N. 1978. *Catastrophe Theory and its Applications*. London: Pitman Publishing Limited.

Prazdny, K. 1984. On the perception of Glass patterns. *Perception*, 13:469–478.

Reiter, R. 1978. On closed world data bases. In *Logic and Data Bases*, H. Gallaire and J. Minker (Eds.), New York, Plenum Press.

Reiter, R. and Mackworth, A.K. 1989. A logical framework for depiction and image interpretation. *Artificial Intelligence*, 41:125–155.

Rock, I. 1983. *The Logic of Perception*. Cambridge: MIT Press.

Sekuler, A.B. and Palmer, S.E. 1992. Perception of partly occluded objects: A microgenetic analysis. *Journal of Experimental Psychology: Human Perception and Performance*, 121(1):95–111.

Sekuler, A.B., Palmer, S.E., and Flynn, C. 1994. Local and global processes in visual completion. *Psychological Science*, 5(5):260–267.

Shashua, A. and Ullman, S. 1988. Structural saliency: The detection of globally salient structures using a locally connected network. In *Proceedings of the Second International Conference on Computer Vision*, Tampa, FL, pp. 321–327.

Shi, J. and Malik, J. 2000. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. (forthcoming).

Smits, J.T. and Vos, P.G. 1987. The perception of continuous curves in dot stimuli. *Perception*, 16(1):121–131.

Smits, J.T.S. and Vos, P.G. 1986. A model for the perception of curves in dot figures: The role of local salience of "virtual lines." *Biological Cybernetics*, 16:407–416.

Stevens, K.A. 1978. Computation of locally parallel structure. *Biological Cybernetics*, 29:19–28.

Stevens, K.A. and Brookes, A. 1987. Detecting structure by symbolic constructions on tokens. *Computer Vision, Graphics, and Image Processing*, 37:238–260.

Takeichi, H., Nakazawa, H., Murakami, I., and Shimojo, S. 1995. The theory of the curvature-constraint line for amodal completion. *Perception*, 24(3):373–389.

van Lier, R.J., Leeuwenberg, E.L.J., and Van der Helm, P.A. 1995. Multiple completions primed by occlusion patterns. *Perception*, 24:727–740.

Wagemans, J. 1993. Skewed symmetry: A nonaccidental property used to perceive visual forms. *Journal of Experimental Psychology: Human Perception and Performance*, 19(2):364–380.

Wagemans, J., Gool, L. van, Swinnen, V., and Horebeek, J. van. 1993. Higher-order structure in regularity detection. *Vision Research*, 33(8):1067–1088.

Williams, L. and Jacobs, D.W. 1997. Stochastic completion fields: A neural model of illusory contour shape and salience. *Neural Computation*, 9(4):837–858.

Williams, L.R. and Thornber, K.K. 1999. A comparison of measures for detecting natural shapes in cluttered backgrounds. *International Journal of Computer Vision*, 34(2/3):81–96.

Witkin, A.P. and Tenenbaum, J.M. 1983. On the role of structure in vision. In *Human and Machine Vision*, J. Beck, B. Hope, and A. Rosenfeld (Eds.), Academic Press, pp. 481–543.

Zucker, S.W. 1985. Early orientation selection: Tangent fields and the dimensionality of their support. *Computer Vision, Graphics, and Image Processing*, 32:74–103.

Zucker, S.W., Stevens, K.A., and Sander, P. 1983. The relation between proximity and brightness similarity in dot patterns. *Perception & Psychophysics*, 34(6):513–522.