Integration, Inference, and Management of Spatial Information Using Bayesian Networks: Perceptual Organization

Sudeep Sarkar, Student Member, IEEE, and Kim L. Boyer, Member, IEEE

Index Terms—Bayesian networks, inference, integration, management of spatial information, perceptual organization.

Abstract—The use of knowledge bases has been advocated by many researchers to make computer vision more stable and reliable. The formalism of Bayesian networks provides a very elegant solution, in a probabilistic framework, to the problem of integrating top-down and bottom-up visual processes, as well serving as a knowledge base. We modify the formalism to handle spatial data and thus extend the application of Bayesian networks to visual processing. We call the modified form the perceptual inference network (PIN). We present the theoretical background of a PIN and demonstrate its viability in the context of perceptual organization. Perceptual organization imparts robustness, efficiency, and a qualitative and holistic nature to vision. Thus far, the approaches to the problem of perceptual organization have been purely bottom up, without much top-down knowledge-base influence, and are therefore entirely dependent on the inputs, which are obviously imperfect. The knowledge base, besides coping with such input imperfection, also allows us to integrate multiple organizations and form a composite organization hypothesis. The PIN imparts an active inferential and integrating nature to perceptual organization in an elegant probabilistic framework.

I. INTRODUCTION

"Perception is not a mere passive recording of information impressed upon my sensory organs by the environment. Rather, it consists of an active construction by means of which sensory data are selected, analyzed, and integrated with properties not directly noticeable but only hypothesized, deduced, or anticipated, according to available information and intellectual capacities."

Gaetano Kanizsa

The motivation for perceptual organization of low-level features, such as edges or regions, into higher level geometric features exhibiting good form or structure (Prägnanz\(^1\)) stems from the observation that ordered geometric arrangements are highly improbable to have arisen by chance and are very likely to be mutually dependent. The inference of such structures reduces the combinatorics of high-level visual processes such as model-based search and structural stereo. Perceptual organization also imparts the robust, qualitative, and holistic nature of human perception to the frail, quantitative, and local character of some of the current algorithms in computer vision [1]. Hereafter, the approaches to perceptual organization in the computer vision community have been mainly bottom up, except for [2], where there is some provision for higher level processing to affect organization. It is our thesis that perceptual organization has two components: One is a purely bottom-up process of recognizing regularities in the image, whereas the other involves the inference (deductive or abductive) of various geometric structures from multiple cues. This viewpoint is also shared by researchers in human visual perception [3].

In this exposition, we develop a model for perceptual organization in the context of machine vision having explicit knowledge about various Euclidean geometric structures. According to our understanding, the use of an explicit knowledge base in perceptual organization is rare if not unique. We believe that knowledge about geometric structures will enable us to organize our external stimuli in a better fashion. We do not claim that our system models the human perceptual system. Our main contribution is the formalism based on Bayesian networks [5] for geometric knowledge-base representation. The knowledge base allows us to infer missing data and gain immunity against noise processes. We propose an efficient methodology, which is based on sound probabilistic semantics and is amenable to multiprocessing hardware, to reason over spatial data. To our knowledge, this is the first application of Bayesian networks to intermediate and low-level vision. As we shall see in the next section, there have been few applications of Bayesian networks to high-level vision. The network formalism also enables us to provide an efficient framework for the integration of multiple sources of information, such as corner detectors, straight line detectors, circle detectors, and ribbon detectors.

\(^1\)Prägnanz is the tendency of a process to realize the most regular, ordered, stable, balanced state possible in a given situation. This is also known as the principle of good form.
In the next section, we review prior work relevant to our approach and identify the relation to ours. In Section III, we provide an overview of our perceptual organization system. Section IV reviews, for the sake of completeness, the work already done by us on preattentive perceptual organization. We propose the new network formalism in Section V. The ideas of the formalism are clarified through an example network in Section VI. Section VII evaluates the computational complexity of the network formalism. Results on real images are presented in Section VIII.

II. RELATION TO PRIOR WORK

Marr [5] was the first to suggest incorporating groupings based on curvilinearity into larger structures in the primal sketch. Witkin and Tenenbaum [1], in their structure-based vision paradigm, recognize the broad implications of perceptual organization for computational theories in machine vision. Lowe [6], [7] demonstrated such computational complexity reduction implications in 3-D object recognition by using simple organizations such as parallelness and collinearity.

The term perceptual organization covers a range of processes including detecting edges, dot grouping [8], and region segmentation [9] to grouping higher level primitives such as in curve decomposition [10], organizing ellipses [11], finding elongated regions [12], finding line organizations for motion sequence matching [13], texture grouping [14], detecting occlusions [15], detecting ribbons to form object hypotheses [16], and fitting superquadrics for form perception [17]. Recent years have seen the increased use of energy minimization criteria in perceptual organization. McCafferty [2] formulated the grouping problem in perceptual organization as an energy minimization problem, which he then solved using simulated annealing. Detecting occluding regions using energy minimization is discussed by Nitzberg and Mumford [18]. Trytten and Tuceryan [19] described an algorithm based on energy minimization to perform curvilinear groupings of incomplete edge contours or edges.

The hierarchical nature of organization has been reiterated by many researchers [20]–[22]. Higher level, complex organizations are gradually built from low-level, less complex features, enabling a better computational control of the process. There have also been approaches to perceptual organization based on some network formalisms. However, none of these approaches is based on Bayesian networks nor do they include probabilistic semantics. Sha’ashua and Ullman [23] compute salient structure from local characteristics using a locally connected network. The output is a saliency map, which is a representation emphasizing salient locations. Mohan and Nevatia [24] use perceptual organization concepts in detecting and describing buildings in aerial images. They recognize the usefulness of the structural relationships made explicit by perceptual organization in complex image understanding. All reasonable feature groupings are first detected, and then, the promising ones are selected by a constraint satisfaction network. The results are extended to curved segments in [25].

The use of Bayesian networks is not new. Binford et al. [26] and Chelberg [27] each use knowledge bases encoded in a Bayesian net to recognize objects in range images in the SUCCESSOR system. Their work differs from ours in two ways. First, theirs is a high-level system that infers complete objects and therefore has a network of parts corresponding to each object. Second, we do not agree with the structure of the Bayesian network as used in [27]. Chelberg suggests that an object causes its parts; therefore, the links should be directed from the objects to their parts. This leads to the conclusion that the parts are conditionally independent, given the object or, more precisely, that the existence of each object part is a conditionally independent event, given the existence of the entire object. For example, this would suggest that given the observation of a table, the observation of each leg and the top would be (five) statistically independent events. We do not agree.


III. SYSTEM OVERVIEW

The aim of our computational paradigm is to develop an algorithm to organize features into highly plausible sets of higher level geometric features that are present in images of objects belonging to a large number of domains. Our organizational philosophy (see Fig. 1) is hierarchical with complex organizations being formed from simple organizations. Coherent global structures gradually emerge from local features. A group of tokens at a particular level is represented by a single more complex structure, along with its emergent properties, at the next level using a fast and efficient organizational scheme. A typical hierarchy would start with constant curvature segments and build the level of tokens, parallelograms, ellipses, and circles. The latter organizations would in turn be organized into regular arrangements of ribbons, ellipses, and parallelograms.

Each level of the hierarchy is constructed using voting methods, graph operations, and knowledge-based reasoning in a new extension of the Bayesian network we call the perceptual inference network (PIN). Regularities in the image tokens based on the Gestaltic principles2 of proximity, similarity, smooth continuity, and closure are detected by the voting methods effecting a search procedure among the image tokens and are shown to be superior to conventional techniques. The Gestaltic associations among tokens are represented by a set of Gestalt graphs, which lets us apply sophisticated graph theoretic techniques. The knowledge base, which is encoded in the form of a PIN, helps us to go beyond the data to predict undetected features and to integrate multiple sources of information. For example, a rectangle has the property that it is closed, it is made up of two sets of parallel lines, and it

2Gestalt is a German word that roughly translates to "organized structure." Gestalt theory is a very general psychological theory that can be used to study and understand aspects of human behavior and experience. Visual perception is particularly suitable for demonstrating actions of Gestalt principles. For a good treatment of Gestaltic perception, see [3].
has right angled corners. These properties, which are detected in a preattentive phase, provide cues for the presence of a rectangle, and the network then enables us to hypothesize its presence, even in the absence of some of the above features, and to generate expectations for such.

Analogous to theories in human vision, our strategy divides broadly into two parts: detecting regularities and similarities in the tokens (preattentive vision) and reasoning (based on a knowledge base built from past experience) to enable one to go beyond the information provided (attentive vision). The voting method provides organizations based on Gestalt principles and the network reasons on those organizations to extract geometric features. The two steps of voting and evidential reasoning are repeated. The type of organizations considered at each level gets increasingly complicated. After two or three levels, we can consider generic part descriptions; therefore, the role of voting gradually diminishes and is replaced by more reasoning on sets of features. Thus, there is a gradual transition from intermediate- to high-level vision processes.

In the next two sections, we describe the preattentive and attentive perceptual organization modules, respectively. We have proposed and defended the preattentive module elsewhere [31]. However, for the sake of completeness, we describe the module briefly in the next section.

IV. PREATTENTIVE ORGANIZATION

The preattentive perceptual organization module consists of a search component implemented using voting methods and a structure-extraction component based on graph theory. The details of the components and their respective design criteria can be found in [31]. The description here is necessarily terse.

A. Graph Theoretic Operations

The basic, single-level preattentive algorithm is depicted in Fig. 2. The Gestalt laws of organization suggest a few basic forms, which we will use throughout the hierarchy to describe the current organization. For example, a ribbon is a symmetric arrangement of proximal tokens sharing a common region. In accordance with the Gestalt laws, we consider five basic graphs: proximity, end point nearness, continuity, similarity, and common region. We call these the Gestalt graphs. The nodes of each graph represent the tokens from the previous level, and the arcs (which may be weighted) denote the existence of the various relationships. The proximity graph is a graph whose links join nodes representing tokens having points close together with similar orientation. The nodes of the end point nearness graph represent the endpoints of the edge tokens, and the links join endpoints that are close or that are parts of the same contour. The links of the continuity graph connect nodes (representing the edge tokens) that are possible continuations based on orientation and intersect or on curvature and center. The similarity graph encodes the fact that two tokens share common photometric or geometric properties. The common region graph is a graph of tokens sharing a common region in the image and is constructed for the first level only.

Fig. 2 also depicts how the various organizations are calculated. The proximity and the common region graphs are “anded” to construct a graph from which parallel structures such as ribbons and mergings are extracted. The links of the resulting graph join tokens having portions of segments that are close together, are parallel, and share a common region. The links of the endpoint nearness graph join line\(^3\) segment endpoints. Each link is quantified by the Euclidean distance between the endpoints, if they correspond to endpoints of different segments, or a negative number, if they are from the same segment. The minimal spanning tree of this graph defines a neighborhood structure [32], and the corresponding fundamental cycles form closed figure hypotheses. The short-
est paths between nodes of valency one, i.e., free endpoints, form strand hypotheses.

Connected components of the graph constructed from the endpoint nearness graph by deleting all links corresponding to the edge segments form junction hypotheses. Intersections between segments are detected by writing out the line segments on an array and identifying crossings. Intersections of the virtual lines (the symmetry axes) representing ribbons give various quadrilateral hypotheses. Identification of continuous lines is done in the following manner. We "AND" the continuity, endpoint nearness, and common region graph to form a graph of tokens representing the relationship that the tokens share a common region, have similar orientation or curvature and center, and their endpoints are close. Complete subgraphs are formed from the connected components of this graph and are "AND"ed with the continuity graph. Cliques in the resulting graph form continuous line hypotheses, such as continuous straight lines, co-circular arcs, and circles.

B. Voting Scheme

The Gestalt graphs defined in the previous section can be constructed using a variety of procedures. Our problem is to find all pairs of tokens satisfying some compatibility relation \( R \), defining the links of the graph. A brute force search over all token point pairs is obviously impractical. Instead, we build global consensus over the image using a voting scheme. Let us say that the compatibility relation is a function of the token attributes \( a_1, a_2, \ldots, a_N \). We consider an \( N \)-dimensional space spanned by these attributes, where each marked point defines a token with its attributes determining its coordinates. In this parameter space, each token votes for all points satisfying the compatibility relation \( R \). This defines a \textit{compatibility region} centered on the token whose shape is dictated by the compatibility relation. Each vote is tagged with the token number. If two tokens vote for any common location in parameter space, then those two tokens will be associated. Note that the voting of the tokens can be done in parallel. For a detailed analysis of the voting method in terms of quantization error, complexity, and its difference from the Hough transform, see [31].

V. ATTENTIVE ORGANIZATION

By attentive organization, we refer to the active process of detecting structure by inference, hypothesis, and confirmation. The process involves a knowledge base of basic information encoded in the form of a Bayesian network. For perceptual organization, we feel a knowledge base aware of geometric figures such as circles, ellipses, rectangles, polygons, and the like is appropriate. A geometric feature can be constructed from various cues, such as combinations of vertices and sides; the whole feature is then inferred from the parts detected. This enables us to detect features in the presence of noise and occlusion.

Bayesian networks, which are also known as belief networks, influence networks, or causal networks, are directed acyclic graphs with nodes representing propositions (or random variables) and arcs signifying direct dependencies as quantified by conditional probabilities.\(^4\) An example Bayesian network is shown in Fig. 3. Bayesian networks do not assume independence among features; rather, they encode the dependencies among features. This can be seen by comparing the factoring of the joint density with the structure of the network in the figure. Consider the following simple example of two lines. Two distinct lines \( l_1 \) and \( l_2 \) are independent. However, they become dependent once we know that they are parallel because that suggests that they may share a common underlying cause. Mathematically, denoting the event of parallelism as \( I \), we have \( P(l_1, l_2) = P(l_1)P(l_2) \) but \( P(l_1, l_2 | I) \neq P(l_1 | I)P(l_2 | I) \).\(^5\) This information is encoded in a Bayesian net by three nodes; two nodes represent the lines, and the third node represents the parallelism relation; links are drawn from the line nodes to the parallel relation node.

Extension: The Perceptual Inference Network

Our problem is to integrate information about various spatial features to form composite hypotheses. Let us say we have \( m \) features \( f_1, \ldots, f_m \) and \( n \) locations \( l_1, \ldots, l_n \). Let us form our random variables as \( I_{ij} \) taking two values: 0 (false) and 1 (true). The event \( \{I_{ij} = 1\} \) denotes the fact that feature \( f_j \) occurs at location \( l_i \). Our aim is to formulate an efficient means of updating the probabilities of these random variables based on the evidence. Given a set of features at some locations, expectations for other features at different locations are formed. We want to devise an efficient means to do so. In total, we have \( mn \) random variables. One way of updating is by specifying a joint probability distribution of \( mn \) random variables, which is very difficult if not impossible.

We make the problem tractable by exploiting the conditional independencies inherent among the variables. Features that are dependent tend to be close together spatially. In the context of a hierarchical system, this assumption is generally true. Dependencies among distant features are captured at higher levels

\(^4\) Although the links are directed, probabilistic messages are passed both ways. The message in the direction of the link is termed \( x \), and the other way is \( \lambda \). \( \lambda \) and \( \pi \) denote the bottom-up and top-down components, respectively [4].

\(^5\) It is interesting to note that in the theory of structural descriptions of models [33], it is assumed that the relations are conditionally independent given the primitives and not the other way around. Our assumption is consistent with this.
of the hierarchy. The local dependency structure will make
the connections among the random variables sparse. Let us say we
can construct a directed acyclic graph among the \( mn \) variables
with the links quantified by the conditional probabilities among
the variables. This will form the Bayesian network that we
can update using the message passing discussed in [4]. The
disadvantage of this solution is its high structural complexity.
We have \( mn \) nodes and typically \( mn-1 \) connections. Yet,
many nodes will not be used because there will be no evidence
to support the feature at the location represented by the node.
We will take advantage of this sparseness.

The sparse nature of the above Bayesian network can be
exploited to reduce the structural complexity further. The
sparseness arises because some random variables are inde-
pendent from another or conditionally independent. Besides
being sparse, the network will also have a number of similar
substructures because the same feature (type) at different
locations will have similar neighborhood structure. We can
group together some of the random variables representing
the same features at various locations. These random variables
share the common factor that they are the same feature
\( f_j \) but in different locations. Hence, they will also share a
common neighborhood structure (in the network) with other
such groupings. The group of random variables \( l_{i1}^f, \ldots, l_{ik}^f \)
form a composite random variable representing the events
that feature \( f_j \) occurs at \( l_i \) through \( l_k \). The grouping will
certainly reduce the number of nodes of the network, but the
computational complexity of each node increases. However,
the increase is less than the reduction in structural complexity,
as we shall see. Note that we might not be able to group
all the random variables representing a feature. For example,
the grouping of \( l_{i1}^f, \ldots, l_{ik}^f \) might not be possible if it leads
to the formation of a cycle in the undirected graph structure.
This concept will become clearer when we discuss a concrete
example.

Grouping the nodes will create various other issues, such as
message conflict resolution, updating of probabilities, and the
construction of conditional probability matrices. We address
these issues next. We refer to the Bayesian net with these
incorporated features as a PIN (see Fig. 4).

A. Belief Revision Using Modified Message Passing

Each composite node (Fig. 5) represents a collection of
random variables denoting the presence of a feature at a
set of locations. The common characteristic is the type of feature
each random variable represents. They differ only in the
location of the feature. We have to specify the conditional
probability of a feature at a particular location given the
parents at compatible locations. The original formulation of
message passing in the Bayesian network will suffice for this
case. However, we need to check for compatible locations
when the message arrives. The conditional probabilities that
need to be specified are of the following form:

\[
P(l_{i1}^f, l_{i1}^u, \ldots, l_{im}^u) \quad \text{for} \ i, l = 1, \ldots, n
\]

where \( l_{i1}^f, l_{im}^u \) denotes the locations of the features \( X \)
and \( U_m \) at \( l_i \), and \( l_{im} \), respectively. We use upper case to refer to
random quantities and lower case to refer to specific values.

The specification of the conditional probability can be
further simplified if we assume that the conditional probability
can be factored into two functions \( g \) and \( Comp \), where one
depends only on the feature type, and the other is a locational
compatibility function:

\[
P(l_{i1}^f, l_{i1}^u, \ldots, l_{im}^u) = g(x|u_1, \ldots, u_m)Comp(l_x, \{u_1, \ldots, u_m\}). \tag{1}
\]

The function \( g \) is a conditional function that expresses the
degree of belief in a feature \( X = x \), given the existence of
other features \( U_1 = u_1, \ldots, U_m = u_m \). This, in general, will
not be a probability function because it is multiplied by the
\( Comp \) function to form the conditional probabilities. For lack
of a better name, we call \( g \) the conditional belief function.
This function is independent of the locational information and
is affected only by the definition of the features. Thus, the
feature “triangle” depends on three lines and three vertices.
We can assert a degree of confidence in the existence of
the triangle based solely on the presence of subsets of the
lines or vertices and the attributes of the formed triangle.
The locational compatibility of line or point features with a
particular triangle is captured by the \( Comp \) function.

The form of the \( Comp \) function is a matter of choice and
depends on the feature sought. The range of the function
should be chosen such that \( P \) is a conditional probability
function. The support of \( Comp \), that is, the domain over
which it is nonzero, depends on the feature represented by
the composite node. We choose a binary valued function that
is unity (true) when the locations of the features lie within a
locational compatibility tolerance. The choice of the tolerance
is an open issue, depending on the amount of positional
error one can accept. The belief in the features at locations
not compatible with the incoming message locations (where
\( Comp = 0 \), or false) are unchanged.
Assuming the form of the conditional probability to be as in (1) means that instead of storing a set of conditional matrices, we need to store only the much smaller $g$ matrix and $Comp$ function definition. As for the other belief parameters, instead of storing one set, the composite node keeps lists of belief parameters for each location separately. Since the beliefs in features at locations not compatible with a particular location are not changed, we can decompose the processing at the node into two modules (see Fig. 5). The first module computes spatial compatibility and chooses the belief parameters of the compatible features for updating. Spatial compatibility can be computed using a function, a look-up table, or an associative memory.

If there is no match for a message, then a new entry is created in the table, obviating the spawning of a new node, as would be required in a brute-force Bayesian network implementation. The flow diagram is shown in Fig. 6. The belief updating is done in the second module by a shared computational resource [4]. To prevent multiple updates from the same direct evidence, each message is also tagged with the identity (ID) of the elementary evidence on which it is based. Thus, duplicate confirmations are avoided by checking the evidence ID at the composite node, as well as the location. This computational resource is time shared by the features at different locations and is shielded from the network by the compatibility computation module (again, see Fig. 5), unlike previous formulations of Bayesian networks.

**B. Network Instantiation**

The composite nodes and associated links of the PIN are instantiated a priori. The root composite nodes are given equal prior probabilities. At the start, each of the composite nodes has a NULL instantiated location table. As new evidence is introduced into the PIN, locational entries are made at the composite nodes corresponding to the hypothesized structure.

**C. Prediction of Structures**

As mentioned earlier, on receiving a message from a neighbor, the locational module of a composite node generates a set of compatible locations. If a feature already exists at the predicted location, its associated probabilities are updated. Otherwise, a new locational entry is created. These locational entries form predictions of the structures corresponding to the respective composite nodes. Although we may lack direct evidence for a given structure, it can often be inferred from messages received at the corresponding composite node. The algorithm to generate compatible locations depends on the feature type. This idea will become clearer with an example PIN in Section VI-B.

**D. Management of Expensive Computational Resources**

Many (most) machine vision algorithms, such as symmetry detection and recovery of shape from shading, are very expensive computationally. It is to our advantage to apply these algorithms opportunistically only when there is promise of success or a need to resolve ambiguity. The formalism of this paper supports this. The evidence for one feature at a particular location may also raise expectations for other features at other locations. A utility maximization module can decide, based on the resulting probabilities, which computational modules (if any) can offer the maximum advantage. This is an open research issue, but the selection criteria clearly should be based on the expected cost of invoking the module and the potential information gain to be realized from it.

**VI. AN EXAMPLE PIN FOR PERCEPTUAL ORGANIZATION**

The PIN formalism, as presented, is general and can be used to reason and integrate spatial information in a variety of problem domains. The choice of features over which to reason depends on the domain in question. In the case of perceptual organization as implemented here, this depends on the organizations of interest, which vary with level within the organizational hierarchy. The first level of the hierarchy consists of constant curvature segments. The second level consists of ribbons, closed figures, strands, and parallelograms. The third level includes parallel ribbons, strands of ribbons, closed cycles of ribbons, regular arrangements of polygons, and intersections. To demonstrate the viability of this approach,
we will concentrate on the second level of the hierarchy with PIN composite nodes representing closure, segments, strands, ellipses, circles, ribbons, parallels, and corners.

A. Structuring the PIN

Given the nodes of the network, we have to decide on the links in the network. If we had correlation data between the random variables representing our nodes, we could have used the algorithm presented in [4]. However, because of the lack of such information and the absence of a comprehensive theory of Bayesian network construction for visual tasks, we relied on heuristic means and manually constructed our PIN. In the process of the construction, we used the following heuristics:

- The direct dependencies among features are represented by links going from nodes representing less organized features to those representing features having more organization. A set of features of less structure making up a feature of more structure is said to cause the feature of more structure. This encodes the fact that the constituent features are dependent once the more organized feature is detected.
- The distance (in terms of links) between two nodes of the PIN representing two geometric figures should be low if the figures are very similar. For example, a rectangle node and a trapezoid node should be closer than a rectangle node and a circle node.
- Any cycles in the PIN are broken by introducing pseudonodes representing intermediate organizations, that is, we intentionally introduce a bit of node redundancy to avoid cycles. We show examples of this below.

An example Bayesian network is shown pictorially in Fig. 7. The network has 23 nodes, where each represents a feature type, as depicted in the figure. We now point out some key aspects of the network structure.

- Node N23 denotes the concept of a parallelogram, which is formed from the concept of a trapezoid and the attribute that the “nonparallel” sides of the trapezoid are parallel.
- A trapezoid N21 is a quadrilateral with a pair of the sides parallel and represented by nodes N19 and N20, respectively.
- A quadrilateral is a polygon (node N15) having four sides. A polygon with three sides is a triangle (node N16). An equilateral triangle (node N18) is a triangle with appropriate symmetry (node N17).
- The feature of closure (node N3) with corners at particular locations (node N14) forms polygonal hypotheses.
- The features of ellipse (node N6) and circles (node N8) are formed from appropriate symmetry (nodes N5 and N7), respectively.

Note that we have three corner nodes (N10, N11, N14) corresponding to different spatial groupings. Placing them into a single node would render the network cyclic. In addition, note that the features represented by N4, N9, and N15 could have been formed from the concepts of strand (N1) and constant curvature segments (N2) directly, bypassing node N3. However, the network would then have cycles. Introducing N3 keeps the network acyclic.

B. Location Compatibility Functions, Comp

As we saw before, each node of the perceptual inference network has a locational compatibility function, Comp, which is defined to calculate the compatibility of a message from a child or parent with its own location set (see (1)). If the location is instantiated, then the belief parameters associated with that location are updated. Otherwise, the new location is
entered in the list of present locations. To check for compatibility, each node generates a set of compatible locations based on the spatial information sent by the parent or child. It then does a point-by-point locational match to ascertain how far the new hypothesized location is from previously noted locations. Compatibility is ascertained based on a proximity tolerance. For our experiments, this is 5 pixels. The choice of this value depends on the spatial resolution of the system. The robustness of the choice is studied in Section VIII-B-1.

Before going into the details of the matching process, we briefly discuss the format of the locational information at each node. This format is peculiar to this implementation and is certainly not general; one can choose any suitable format. Fig. 8 gives abstract representations of the data structures that store the locations of points and attributes characterizing a particular feature. The spatial information of the strands (N1), constant curvature segments (N2), and the closed tokens (N3) are stored pointwise. Only the best fitting circular arc and/or straight line endpoints are stored for the “all curves” closed figure (N4), the triaxial curve straight line closed figure (N9), and the polygon (N15). Similar structures are stored for the other features. The ellipse node (N6) and circle node (N8) store the relevant ellipse and circle parameters. Computing compatible locations for nodes like the triangle, the trapezoid, or the parallelogram nodes is straightforward; we just match the endpoints. For N6 and N8, we consider the best fitting ellipse or circle, respectively. Nodes N4, N9, and N15 deserve special attention. For a message received at N4 from N6, we segment the ellipse at those points where the latus rectum intersects the boundary. This generates the constant curvature segments for node N4. To generate compatible locations for spatial information from node N3 to N4, N9, and N15, we proceed as follows. The locational information for node N3 is stored pointwise and segment into constant curvature segments as detected by the low-level contour segmentation algorithm [34]. We need to repeat the segmentation of each closed chain of pixels into constant curvature parts to generate the appropriate set of compatible arcs or straight lines.

The process of calculating compatible locations for N4, N9, and N15 for a message from N3 is depicted in Fig. 9. To begin, we compute the possible intersection points of the arcs or straight lines, for example I1, I2, I3, and I4 in Fig. 9. Arcs and straight lines connecting a subset of intersections will form our best decomposition into arcs or straight lines. Thus, for example, a better description of the closed figure in Fig. 9 may be arcs connecting I4 to I1, I1 to I2, and I2 to I4 without involving I3. To choose this optimal subset, we transform the problem into the shortest-path problem in graph theory. The set of possible intersections form the nodes in a graph whose links denote the best fitting arc or line between the intersection points. The weight of a link from node i to j, fij is the fit error considering the part of the cycle between intersections i and j. Note that fij ≠ fji, because they represent the directed fit error between i and j and therefore involve different points. To choose the best decomposition, we select the cycle that starts at a point and ends at that same point having the best fit index. The fit index is defined as the total fit error divided by the number of intersection points constituting that cycle. We search for the optimal cycle by listing the shortest paths starting and ending at a common node and choose the one having the minimum fit index. We use the standard graph algorithm of Floyd and Warshall [35], based on dynamic programming, to calculate all shortest paths between all pairs of nodes.

The optimal segmentation found above forms our compatible location for a message from N3 to nodes N4, N9, or N15. The algorithm for the three nodes N4, N9, and N15 is the same except for the types of primitives we fit between the intersection points to set up our graph. For node N4, we consider circular arcs. For node N9, we consider both the circular and straight lines and choose the one having the minimum error. We compute the compatible locations for node N15 by considering straight line fits between points. To compute compatible locations for the quadrilateral node (N19), we consider the four largest lines of the polygons as constituting the quadrilateral boundary for messages coming from node N15. Similarly, we only consider the three largest sides of the polygon for the triangle node.

C. Conditional Belief Functions g

The conditional belief function g (see (1)) captures the
belief in the specified feature type, given the status of the corresponding parent nodes. The conditional probabilities of the underlying Bayesian network are partly characterized by $g$ and partly by the locational compatibility function $Comp$. 

Fig. 10. Organizations detected by the preattentive perceptual organization module: (a) Gray-level image; (b) edge contours; (c) constant curvature segments; (d) closed boundaries; (e) edge strands; (f) parallel edge segments.
Fig. 11. Organizations detected using the attentive module on the aerial image. (a) New closed boundary hypotheses; (b) parallelograms; (c) ellipses; (d) circles; (e) ribbons; (f) corners.
which was discussed before. We consider the following general form for the conditional belief functions:

\[ g(X = 1|u_1, \ldots, u_n) = C(u_1, \ldots, u_n) \times (1 - e^{-f_{\text{geom}}}) \]  

(2)

where \( X \) is the composite node under consideration, \( u_1, \ldots, u_n \) are the values taken by the random variables represented by \( X \)'s parent composite nodes, and \( C(u_1, \ldots, u_n) \) is a value that depends on the state of the parents. We found from experimentation that the absolute value of the constant is less important than the relative values of the constant for different parent states. The second factor captures the geometric information in the primitive. The exponential is chosen to limit this factor to the interval \((0, 1)\). \( f_{\text{geom}} \) is a factor based on the size of the features and some defining relations, like parallelism for a parallelogram. Qualitatively speaking, the larger the feature, or the more strictly it satisfies the defining relation of a feature type, the larger the value of the belief. Other functional forms with this qualitative behavior may also suffice. In Section VIII-B, we investigate the robustness of the form of the conditional probability function.

In the present implementation \( f_{\text{geom}} \) is calculated in the following manner. For the nodes representing some form of closed convex figure \( (N3, N4, N9, N13, N15, N16, N18, \text{and} \ N19) \), we use the following form:

\[ f_{\text{geom}} = \frac{(\prod_{i=1}^{n} l_i)^{1/n}}{l_{av}} (\xi_{\text{convex}}) \]  

(3)

where \( l_i \)'s are the length of the \( n \) constituting segments of the closed figure, \( l_{av} \) is a normalizing constant and is chosen to be the average length of segments in the image, and \( \xi_{\text{convex}} \) is 1 if the figure is convex and 0 if not. Clearly, it would be a useful enhancement to measure the degree of convexity on a continuum of some sort, but this form suffices for the present. The geometric mean in the numerator is chosen to suppress the spurious generation of hypotheses like a quadrilateral with one side whose length is zero, which we noted were formed with an arithmetic average. The trapezoid node \( N21 \) and the parallelogram node \( N23 \) has the same expression for \( f_{\text{geom}} \) as that above with extra multiplicative terms of the form \( \cos^2(\theta_{\text{diff}}) \), where \( \theta_{\text{diff}} \) is the angular difference between lines, to penalize for nonparallelness in the constituting sides. For the ellipse node \( N6 \), we choose \( f_{\text{geom}} = |b/a - 1| \sqrt{\text{Area}/l_{av}} \), where \( b \) and \( a \) are the minor and major axis, respectively. Circle node \( N8 \) has \( f_{\text{geom}} = \sqrt{\text{Area}/l_{av}} \).

D. Prior Probabilities

One of the common arguments against Bayesian probability theory is that we must always specify prior probabilities. In our case, we have to specify the prior probabilities of the root nodes. We do not see the specification of prior probabilities as a disadvantage; rather, it enables one to incorporate already-acquired knowledge in a very concise manner. The prior probabilities, for our example, are the prior probabilities of occurrence of the root node features like strands, constant curvature segments, corners, and symmetry. There may be domains where the occurrence of, say, corners is very low; this knowledge can be very concisely incorporated in the reasoning process through the assignment of estimates of prior probabilities. In the absence of any prior knowledge, we assume equal priors, which is a solution that maximizes the entropy of the distribution. This is the best we can do in such a situation. It is significant that the effect of prior probabilities in a probabilistic system decreases as new evidence is gathered. Thus, minor differences in prior probability assignment do not matter significantly in the long run. This is shown in Section VIII-B.

E. Evidence Instantiation

The pieces of evidence for our case are the organizations detected in the preattentive phase. These include strands, constant curvature segments, parallelograms, and ribbons. Each piece of evidence activates the node belonging to the corresponding feature type. This evidence is virtual in the sense that it conveys a graded degree of belief about the underlying features. The introduction of virtual evidence can be modeled as a dummy node \( Z \) posting a message \( P(Z = z|x) \) to the network node \( X \). The message is the conditional probability of state represented by the evidence given the state of the node \( X \). For our case, this will be an estimate of the probability of the detected feature. This probability can be measured in terms of the photometric and geometric evidence we have for the feature. For a closed boundary, we set the message equal to the ratio of the total length of the segments to the length of the perimeter of the hypothesized closed boundary. For parallels, it is the fraction of overlap. Constant curvature segments are assigned confidence measures according to the straight line or arc fit error they exhibit.

Note that although our evidence is based on geometric properties, this need not be. Since we start with edge segments, which are necessarily geometric entities, and since the process of perceptual organization is primarily geometric, the emphasis is on geometric properties. However, photometric properties may be used to confirm or reject a hypothesis. This will involve the intelligent management of computational resources because the gathering of photometric evidence is generally expensive. As shown in Fig. 1, special-purpose visual algorithms form a different block of our system and are under investigation.

VII. ALGORITHMIC COMPLEXITY

In this section, we analyze the proposed perceptual organization system from the standpoint of algorithmic complexity and present some timing data. Since the preattentive module is analyzed in [31], we simply state the results here. The complexity is computed with respect to the input size, which is the number of tokens \( N \). The preattentive algorithm is depicted in Fig. 2. As we can see, the algorithm consists of two parts: construction of the Gestalt graphs using voting and the extraction of structure using graph theoretic operations. The voting algorithm was shown to have computational complexity of \( O(N^2) \). However, the average case complexity is \( O(N_{\text{assoc}}) \), where \( N_{\text{assoc}} \) is the expected number of associations, which is generally \( O(N) \). Thus, voting
methods are better than brute-force search, which is always \( O(N^2) \).

The graph theoretic operations used are those to AND graphs, find connected components, compute all shortest paths, calculate the minimal spanning tree, and search for maximal cliques. Anding two graphs is an \( O(N^2) \) operation but with a small multiplicative constant. We find connected components by depth-first search, which is \( O(N) \). The Floyd and Warshall algorithm to compute all shortest paths is \( O(N^3) \), and the minimal spanning tree algorithm is \( O(N^2) \). Finding cliques is \( NP \) complete, but we are fortunate in that the clique size in our graph (the one to find continuous segments) is very small: typically 4 to 5 vertices. Thus, the clique finding algorithm is not the bottleneck. In practice, the execution time is dominated by the polynomial all shortest paths algorithm.

The next part of our algorithm is the attentive part implemented in the PIN. Since a PIN is a form of Bayesian network (except it has a locational matching module), we use the complexity results of Bayesian networks. Inference using general Bayesian networks is \( NP \) hard [36]. However, if we use a singly connected network, then the inference time is on the order of the diameter of the network. Assume that we have \( N \) evidence tokens. The instantiation of each token will take constant time to propagate. However, at each node, we have to do a locational match, which is \( O(\log N) \) when efficient data structures, such as quad trees, are used. Thus, taking all the \( N \) tokens together, belief revision will take \( O(N \log N) \) time when using efficient data structures to store the locational information at each node.

### VIII. RESULTS

The perceptual inference network formalism presented in this paper is very powerful, having numerous possible applications. Here, we present results demonstrating the promise and the viability of the network formalism. The results presented should be judged from that aspect.

We implemented the network shown in Fig. 7 from nodes \( N1 \) to \( N23 \) representing knowledge up to the concept of a parallelogram, ribbons, circles, ellipses, polygons, and closed figures. The code was written in C++ and runs on Sun Sparc-IPX. The PIN is instantiated first with a NULL locational list at each composite node. Each composite node is an object, and the associated methods are the message-handling algorithms. Given a gray-level image, we first detect edges using the optimal zero-crossing operator (OZCO) [37]. The edge contours are segmented into constant curvature segments using a modified form of Wuescher and Boyer’s algorithm [34]. The voting and the graph theoretic modules detect various relations like parallelism, curved symmetry, closure, and strands in a purely bottom-up fashion. These relations are introduced as evidence for the appropriate nodes of the PIN, initiating a sequence of message passing. The output is the set of salient hypotheses represented by some nodes of the PIN like that for parallelograms, quadrilaterals, ellipses, circles, and ribbons with associated probabilities.

In this section, we present results on two real images: an aerial image and an indoor scene. On the aerial image, we also demonstrate performance characteristics of the PIN with respect to the chosen parameters.

#### A. Aerial Image

The first set of results are on the aerial image shown in Fig. 10(a). The scene has a number of linear and curvilinear structures of interest, which is typical of aerial scenes. The detected edges are shown in Fig. 10(b). These edges were segmented into constant curvature primitives (shown in Fig. 10(c)) and given as input to the voting and graph theoretic modules. The latter produced organizations such as parallels, ribbons, closures, and strands, where each is associated with a probabilistic degree of confidence. There were 103 constant curvature segments, 47 closures, nine strands, 12 ribbons, and 16 parallels, as shown in Figs. 10(c), (d), (e) and (f).

The relations detected by the preattentive voting and graph theoretic modules are introduced as evidence into the PIN. After the network settles to equilibrium, which takes 72 s on a Sun Sparc-IPX, we have various organization hypotheses with associated probabilities. Since it is impractical to give a complete listing of the node probabilities, we show some of the organizations as images. Fig. 11 depicts a set of such organizations. The complete set of closure hypotheses represented by node \( N3 \) of the PIN are shown in Fig. 11(a). The predicted sets of parallelograms, circles, ellipses, ribbons and corners are shown in Figs. 11(b)-(f). There are 67 closures, 26 parallelograms, 58 ellipses, 58 circles, and 11 ribbons. Note that these are just hypotheses based on the relational evidence detected by the attentive module. Expensive photometric-based custom feature detectors can now be intelligently applied on the hypothesized image regions based on the associated probabilities. Further discussion of these ideas is beyond the scope of this paper.

#### B. Performance Evaluation

On reaching equilibrium, the PIN produces hypotheses of various organizations, where each is associated with a probability value. In this section, we investigate the robustness of output probability values with respect to the chosen system parameters. The PIN is specified by its graphical structure, the conditional probabilities, and the prior probabilities of the root nodes. The theoretical study of the graphical structure is an open research area and is not considered here. The conditional probabilities are specified by the function \( g \) (see (1)) and the locational compatibility function \( Comp \) (see (1)). The function \( g \) (see (2)) in turn is a product of a constant and a functional

<table>
<thead>
<tr>
<th>( n )</th>
<th>Parallelogram (N23)</th>
<th>Quadrilateral (N19)</th>
<th>Ribbon (N13)</th>
<th>Ellipse (N6)</th>
<th>Circle (N8)</th>
<th>Time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>27</td>
<td>27</td>
<td>11</td>
<td>64</td>
<td>64</td>
<td>69</td>
</tr>
<tr>
<td>4</td>
<td>28</td>
<td>28</td>
<td>11</td>
<td>63</td>
<td>63</td>
<td>71</td>
</tr>
<tr>
<td>5</td>
<td>26</td>
<td>26</td>
<td>11</td>
<td>58</td>
<td>58</td>
<td>72</td>
</tr>
<tr>
<td>6</td>
<td>23</td>
<td>23</td>
<td>11</td>
<td>51</td>
<td>51</td>
<td>74</td>
</tr>
<tr>
<td>8</td>
<td>22</td>
<td>22</td>
<td>10</td>
<td>44</td>
<td>44</td>
<td>67</td>
</tr>
<tr>
<td>10</td>
<td>22</td>
<td>22</td>
<td>10</td>
<td>43</td>
<td>43</td>
<td>75</td>
</tr>
<tr>
<td>15</td>
<td>17</td>
<td>17</td>
<td>9</td>
<td>30</td>
<td>30</td>
<td>78</td>
</tr>
</tbody>
</table>

Fig. 12. Table showing the variation of the number of hypotheses of some salient nodes with parameter \( t \) of the \( Comp \) function. The last column lists the execution time for each value of \( t \).
form. We study the robustness of the PIN probability outputs with regard to these constants, functional form, and the Comp function. In addition, we also probe into the necessity of equal priors assumption for the root nodes of the PIN.

1) Locational Compatibility Function Comp: For these experiments, the support of the locational compatibility function (the domain over which its value is one) is $t = 5$ pixels. This choice depends on the necessary spatial resolution. A large value of $t$ means that features that are spatially distant can be declared to be compatible. This reduces the ability to resolve nearby features and tends to produce fewer hypotheses. However, as we shall see, the reduction in the number of hypotheses is not drastic.

For this discussion, we focus on the variation of the number of hypotheses of a few salient nodes that are distributed over the network: $N23$ (parallelogram), $N13$ (ribbon), and $N6$ (ellipse) as $t$ ranges from 3 to 15 pixels. The results are shown in Fig. 12. Note that the variation of the number of hypotheses is low around the chosen value of the parameter ($t = 5$). Only for very high values (like 15) is the number significantly different. However, the execution time, which depends on the number of introduced pieces of evidence, is approximately constant. From these results, we infer that the PIN is robust with respect to the locational tolerance parameter $t$.

2) Prior Probabilities: The prior probabilities specified for the PIN are those of the root node features. In the absence of information to the contrary, we assume equal priors for the root nodes. We now want to consider the robustness of this assumption by comparing the behavior of the final PIN probabilities for different priors.

As an indicator of performance, we will study the behavior of the probability assignments of nodes $N23$ (parallelogram), $N13$ (ribbon), $N6$ (ellipse), and $N8$ (circle). The behavior of the other nodes is similar. Since the random variables are binary, we consider the probability of just one state, namely, that of existence. For each composite node, we plot the probabilities associated with the feature at different locations.
Fig. 15. Robustness of the choice of the constants $C_{1}$ for nine random perturbations of the constants. The features considered are as follows: (a) Parallelogram ($X3$); (b) Ribbon; (X13); (c) Ellipse (X6); (d) Circle (X6).

Table 16. Table of similarity coefficients $C_{sim}$ for nine random perturbations of the constants $C_{1}$ (for prior approximations.

<table>
<thead>
<tr>
<th>Nodes</th>
<th>$C_{sim}$ for 9 random perturbations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parallelogram</td>
<td>0.9961  0.9991  0.9324  0.9995  0.9995  0.9995  0.9995  0.9995  0.9995</td>
</tr>
<tr>
<td>Ribbon</td>
<td>0.9969  1.00  0.9993  0.9994  0.9995  1.0  0.9995  0.9995  0.9995</td>
</tr>
<tr>
<td>Ellipse</td>
<td>0.9785  0.9779  0.7979  0.9227  0.9990  0.9999  0.9992  0.9990  0.9990</td>
</tr>
<tr>
<td>Circle</td>
<td>0.9682  0.9682  0.9682  0.9682  0.9682  0.9682  0.9682  0.9682  0.9682</td>
</tr>
</tbody>
</table>

and see how they change with different priors.

Notice that the relative values of probabilities are far more important than their absolute assignments. Therefore, we use a normalized correlation measure to quantify the similarity between two probability assignments. Let $\{P_{1}^{i}\}_{i=1,\cdots,n}$ denote one set of probability assignments for $n$ locations and $\{P_{2}^{i}\}_{i=1,\cdots,n}$ for the other. The similarity $C_{sim}$ between the two sets is computed by

$$C_{sim} = \frac{\sum_{i=1}^{n} P_{1}^{i} P_{2}^{i}}{||P_{1}|| \cdot ||P_{2}||}$$

where the denominator is the product of the norms of the two probability vectors.

Fig. 17. Functional form of the dependence of the conditional probabilities with $f_{geom}$. The exponential solid plot is the one that is chosen for the PIN. This is compared against the dot-dashed and dotted choices.

We started from nine random assignments of prior probabilities and compared each result with that resulting from equal prior assignment. The results are shown in Figs. 13 and 14. The
Fig. 18. Performance plots depicting the flexibility of specification of the functional dependence of the conditional probabilities with $f_{\text{prom}}$. The features considered are as follows: (a) Parallelogram $(N23)$; (b) Ribbon $(N13)$; (c) Ellipse $(N6)$; (d) Circle $(N8)$.

<table>
<thead>
<tr>
<th>Nodes</th>
<th>$C_{\text{sim}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dotted function</td>
</tr>
<tr>
<td>Parallelogram</td>
<td>0.9923</td>
</tr>
<tr>
<td>Ribbon</td>
<td>0.9983</td>
</tr>
<tr>
<td>Ellipse</td>
<td>0.9856</td>
</tr>
<tr>
<td>Circle</td>
<td>0.9829</td>
</tr>
</tbody>
</table>

Fig. 19. Table of similarity coefficients $C_{\text{sim}}$ for nine random perturbations of the constants $C_{(u_1,\ldots,u_n)}$ compared with the chosen values. The rows are for four different composite nodes of the PIN.

plots of Fig. 13 are of the final probabilities at a composite node versus the feature at different locations. Note that the relative assignments of the probabilities are preserved most of the time for different priors, although the absolute values vary. In other words, it is the shape of the plots that interest us, rather than their vertical positions. This is also validated by the calculated similarity values $C_{\text{sim}}$ for the nine different priors. The consistently high values imply that the equal prior assumption is not constraining.

3) Conditional Probability Constants: One of the components of the conditional probability at each node (see (1)) is the function $g$, which in turn involves (see (2)) a multiplicative constant $C_{(u_1,\ldots,u_n)}$. These constants have been chosen empirically from the interval $[0,1]$, and in this section, we study the robustness of the choice. As a performance evaluation criterion, we again consider the probability assignments of four salient nodes in the network, namely, $N23$ (parallelogram), $N13$ (ribbon), $N6$ (ellipse), and $N8$ (circle). We added uniformly distributed (between -0.25 and 0.25) noise to the chosen values constrained to the valid interval of $[0,1]$ and studied the final probability distributions. The similarity between this final probability distribution is again compared with that for the chosen constants using (4). We ran nine different trials, and the results are shown in Figs. 15 and 16.

The plots of Fig. 15 are of the probability of existence of features versus their location index for different random perturbations. From the plots, we can readily observe the qualitative similarity of the relative probability assignments for different values of the constant $C_{(u_1,\ldots,u_n)}$. This is further reinforced by the quantitative similarity index $C_{\text{sim}}$ shown in Fig. 16, which compares the similarity of the probability assignments for nine noise-added $C$ values with that for the chosen values. Thus, we infer that the results of the PIN are indeed robust against perturbations of the values $C_{(u_1,\ldots,u_n)}$. 
Fig. 20. Organizations detected by the preattentive perceptual organization module: (a) Gray-level image; (b) edge contours detected; (c) constant curvature segments; (d) parallel edge segments; (e) edge strand hypotheses; (f) closed boundary hypotheses.
4) Functional Dependence of the Conditional Probabilities: The definition of $q$ in the conditional probability expression (see (1)) also involves a function involving a geometric term $f_{geom}$ (see (2)). We have chosen an exponential function for our purpose (see solid plot in Fig. 17). The function maps values from $[0, \infty)$ to the range $[0, 1]$. To study the dependence of the PIN on this functional form, we employ the same performance evaluation scheme as before, namely, considering the final probability distribution of some salient nodes for different functional forms. We have considered two different piecewise linear functional forms that saturate at one as shown in Fig. 17. The dot-dashed line and the dotted line define a boundary for other possible monotonic mappings that can be expected to have more amenable behavior than the bounding piecewise linear mappings.

The final probability plots are shown in Fig. 18. The dot-dashed and the dotted plots correspond to the dot-dashed and dotted functions shown in Fig. 17. From the plots, we can see that the qualitative similarity between the relative probabilities is significant, although not so much as we saw in the case of the other parameters. The associated qualitative similarity index $C_{sim}$ is shown in Fig. 19. The similarity values are for the dot-dashed and dotted functions with the chosen exponential function. Note the high values of the index. For the ellipse composite node (N6), the indices seem to vary more (0.9173-0.9856) than previously encountered. Considering the wide range of distortion we have considered on our original function, the performance is quite stable.

C. On the Laboratory Image

The next set of results use an indoor laboratory image.
The same edge detector and constant curvature segmentation algorithm was used. The preattentive organizations are shown in Fig. 20. There are 34 cycles, 31 strands, one curved ribbon, 43 parallels, and 147 constant curvature segments.

Introducing the preattentive evidence into the network and letting it settle yields the organizations depicted in Fig. 21. As a final count, we have 57 parallellograms, 113 closures, 92 ellipses, 103 polygons, and one ribbon. Note that we now have a number of parallellogram hypotheses. We also have a set of ellipses predicted that approximate the parallellograms detected. These structures were not detected in the preattentive phase.

The execution time for this image was 86 s on a Sun Sparc-IPX.

From these results, we can infer that the network formalism is certainly viable. It is able to integrate and hypothesize organizations using partial information from multiple sources. The method is highly parallelizable and can offer clear spatial reasoning semantics.

IX. CONCLUSION

We presented a theoretical foundation for spatial data integration and reasoning based on the extension of Bayesian networks to perceptual inference networks. The PIN retains all the positive characteristics of Bayesian networks such as parallelizability, clear probabilistic semantics, and efficient updating. The PIN formalism was demonstrated in the context of perceptual organization through real examples. The system was demonstrated to have stable response with respect to perturbation of its probabilistic parameters.

In the near term, perceptual organization-based hypothesis generation will greatly ease the computational burdens of model-based object recognition. Among future work is the possible use of the network to control expensive computational resources dynamically. In addition, theoretical work is needed toward automatic tree-structured PIN construction.

REFERENCES

Sudeep Sarkar (S'90) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Kanpur, India, in 1988. In 1990, he received the M.S. degree in electrical engineering from Ohio State University (OSU), Columbus. He is currently a doctoral candidate with the Department of Electrical Engineering at OSU and a University Presidential Fellow. His interests include computer vision, chess, and soccer.

In 1988, he was awarded the general proficiency award for the best graduating student in electrical engineering. He also received the National Talent Scholarship from the Government of India. In 1990, he was awarded the Graduate Research Forum award for his thesis research.

Kim L. Boyer (M'86) received the B.S.E.E. (with distinction), M.S.E.E., and Ph.D. degrees, all in electrical engineering, from Purdue University, West Lafayette, IN, in 1976, 1977, and 1986, respectively. From 1977 to 1981, he was with the Satellite Transmission Laboratory, Bell Telephone Laboratories, Holmdel, NJ, where he specialized in digital echo cancellers and speech interpolation. He received a U.S. patent for his work on extended delay echo cancellers. From 1981 to 1983, he was with the Image Processing Department, COMSAT Laboratories, Clarksburg, MD, where he developed a real-time transform-based image coder and investigated block formation strategies, noise weighting functions, and 2-D window functions for direct 2-D transform coding of composite color video. Since 1986, he has been with the Department of Electrical Engineering at Ohio State University, Columbus, where he is currently an Associate Professor. There, he founded the Signal Analysis and Machine Perception Laboratory (SAMPL). His current research interests include medical image understanding; stereopsis in weakly constrained environments; optimal feature extraction; inexact structural matching; robust methods for perceptual organization in image, range, and flow data; and the organization of large structural databases.

Dr. Boyer has authored or co-authored about 40 technical publications, is Co-Chair of the IEEE Robotics and Automation Society Technical Activities Board Technical Committee on Computer and Robot Vision, is an Associate Editor of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE and is co-chair of the SPIE Conference on Applications of Artificial Intelligence XI: Machine Vision and Robotics. He was the recipient of the 1991 OSU College of Engineering Lumley Research Award.