Uncertainty in Artificial Intelligence 4
R.D. Shachter, T.S. Levitt, L.N. Kanal, J.F. Lemmer (Editors)

The structure of Bayes networks for visual recognition

John Mark Agosta

Robotics Laboratory
Stanford University
Stanford, CA 94305
johnmark@coyote.stanford.edu

I. The problem

This paper[1] is part of a study whose goal is to show the efficiency of using Bayes networks to carry out model based vision calculations. [Binford et al. 1987] Recognition proceeds by drawing up a network model from the object's geometric and functional description that predicts the appearance of an object. Then this network is used to find the object within a photographic image. Many existing and proposed techniques for vision recognition resemble the uncertainty calculations of a Bayes net. In contrast, though, they lack a derivation from first principles, and tend to rely on arbitrary parameters that we hope to avoid by a network model.

The connectedness of the network depends on what independence considerations can be identified in the vision problem. Greater independence leads to easier calculations, at the expense of the net's expressiveness. Once this trade-off is made and the structure of the network is determined, it should be possible to tailor a solution technique for it.

This paper explores the use of a network with multiply connected paths, drawing on both techniques of belief networks [Pearl 86] and influence diagrams. We then demonstrate how one formulation of a multiply connected network can be solved.

II. Nature of the vision problem

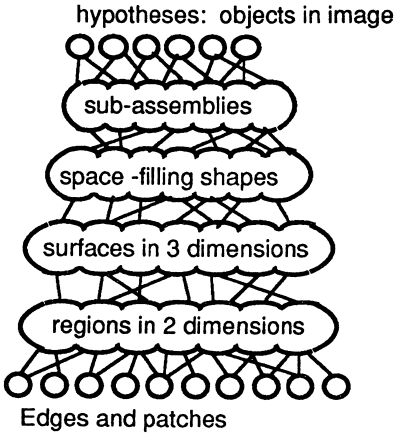
The objects within a visual image offer a rich variety of evidence. The image reveals objects by their surface edges, textures, color, reflectance, and shadows. Save for extreme cases, only a part of this evidence is necessary to recognize an object. The vision problem is no simpler because of the surfeit of evidence: Each kind of evidence presents a new proble.n. The researcher may approach the preponderance of clues by concentrating only on a limited variety, such as those which are easiest to calculate, or lead to the most efficient algorithm. Bayes methods encourage the use of a wider variety of evidence since they traditionally have been developed to integrate diverse and subtle sources of evidence.

The problem of image recognition has close kin. By considering other varieties of evidence, an object could be identified by those not available in a visual image, such as tactile feel or motion when it is disturbed. Similarly there are vision problems for which recognition is not necessary, such as visual obstacle avoidance. Recognition is one aspect of understanding the scene; the scene

[1] This paper grew out of extensive discussions with Tom Binford, Dave Chelberg, Tod Levitt and Wallace Mann. I owe a special debt of gratitude to Tom for introducing me to both the problem and a productive way to approach it. As usual, all errors are the sole responsibility of the author.

also may be analyzed to infer the location of the viewer relative to the object and to make other functional statements. Conceivably the process of recognition might proceed to a higher level recognition of some situation or "Gestalt." These concerns are outside the scope of this paper.

Model based vision consists of two activities; first, modelling the objects to be looked for – the predictive phase, then identifying objects by analyzing a raster image – the inferential phase. Vision models are built from a top down decomposition of an object's geometry into geometric primitives that further decompose into primitive observable features. The model of the object is decomposed into sub-assemblies that are in turn decomposed to obtain relations among volume filling primitives, in our case "generalized cylinders" and their intervening joints. Volumes have surfaces that appear as patches in the image. The patches are projected onto the image visual plane as regions bounded by edges and junctions, the lowest level features in the hierarchy.



Vision recognition based on a model proceeds by grouping image features at a lower levels to identify features at higher levels.The bottom up grouping process is driven by the decomposition model of objects expected to be within the image.

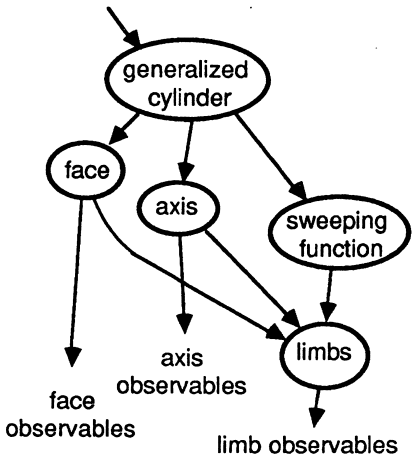
III. Issues in formulation

Grouping is a primary process of recognition, and occurs at each level of the decomposition hierarchy. A predictive arc from a feature to a lower level feature implies both the appearance of a lower level feature and its location relative to the feature. If a predictive arc did not entail some position information, then the network could not perform grouping of features. Consider the contrary case: A network constructed out of complete but non-localized evidence of the number and kind of low level features that compose an object. This is still only weak evidence for the appearance of the object. Unfortunately grouping adds dependencies that complicate the network structure. Here I discuss the formulation of these dependencies.

A. Tree structured hierarchies.

The formulation of a decomposition model implies a probability network where top level constructs predict the appearance of lower level ones. When the observable parts at a stage can be decomposed independently, the network becomes tree structured. The tree is rooted (at the "top", an unfortunately confusing use of terms) in a hypothesis about the appearance of an object. It grows "down" to leaf nodes that represent image primitives. The process of recognition begins when evidence from an image instantiates primitives. Then by inference from lower level nodes to higher levels, the calculation results in the probability of the object hypothesis given the evidence.

For example, a generalized cylinder consists of a face, an axis and a sweeping function for the face along the path of the axis. The face of the generalized cylinder appears independently of the axis. The lower level observables of both face and axis remain independent. In contrast, the sweeping rule and the limbs that it predicts depend on both the face and axis.

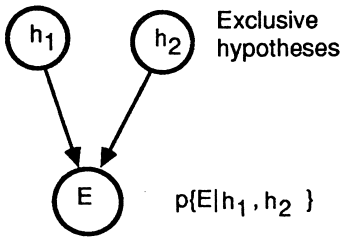


B. Ambiguity

Optical illusions that can appear to be different objects are an example of ambiguity functioning in the human vision system. The novelty of such images indicates their rarity. This is evidence that the eye, in the process of bottom up grouping rarely resorts to backtracking. Ambiguity can be formulated as the existence of a higher level construct that clarifies an ambiguity. more than one level below. It is expressed by influences that skip over levels. They express kinds of arguments that stand in contrast to the grouping-at-each-level kind of reasoning we apply. Hopefully, the intermediate constructs in the hierarchy are rich enough so that they are not necessary.

C. Exclusion

Multiple parents in the network may be used to express exclusion, so that recognizing one object implies the other does not exist.



The evidentiary node for their exclusion is a sibling of both objects. Instantiating *E* offers evidence for one hypothesis but not both. The distribution function of the evidence given the object hypotheses, $p(E \mid h_1,h_2)$, resembles an exclusive-or function. We can demonstrate the strong dependence between object nodes that this evidence generates by flipping the direction of the arc between a hypothesis and the evidence. By Bayes rule this generates an arc between hypotheses. Thus the truth of one hypothesis upon observing the evidence depends strongly on the truth of the other. This characteristic multiple parent structure allows formulation by Bayes networks of conflicting hypotheses sets as presented by Levitt [Levitt 1985]. It is useful for the purposes of formulation to have evidentiary nodes that excludes hypotheses for different objects from evidence the same location.

D. Co-incidence

Just as multiple parents can express exclusion, they can be used to infer two models at the same location. Enforcing co-incident locations for different models could be a useful modeling tool. Imagine an object that could be posed as two separate models depending on the level of detail. For instance the "Michelin Man" could be modeled as both a human figure and as a stack of tires. As evidence for him, the perceptor would expect to find both a man and a stack of tires in the same location.

More common objects may also be composed as a set of co-incident models. For example a prismatic solid may be interpreted differently as generalized cylinders, depending on the choice of major axis. We may infer more than one of these generalized cylinders occupying the same location from which we infer the one object that predicted the set.

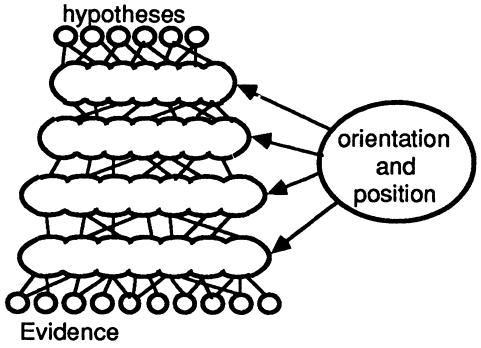
Composition of the same object as several co-incident models is not to be confused with the decomposition hierarchy. Decomposition is essentially a conditional independence argument, that the separate features into which an object is decomposed can be recognized, given the object appears in the view. Composition as several co-incident models, like exclusion formulations, depends on one lower level feature being predicted by multiple higher level features. Such multiple parent structures describe relations among their ancestors. In co-incident models, furthermore, the multiple higher level features are resolved into the one ancestor that admits of co-incident interpretations. This generates a network with multiple paths.

E. Global location and the use of proximity information.

In general, evidence for exclusion and co-incidence is of the kind from which the proximity of two higher level constructs may be inferred. For instance, when an observable feature in the image is due to the joint between two other features, then the proximity of these two may be inferred.

Further, all parts of the hierarchy are influenced by object position, orientation and

articulation. All variables are a function of these global variables. They form a set of variables with universal influences, as shown here:



This generates a set of arcs that violate tree type hierarchies. These dependencies allow information about feature orientation and position to be made available to all superior nodes, once an inference about it can be made from a lower level.

Again a question arises about the trade-off between making this influence explicit, or entailing "softer" proximity information within the network structure. Since the geometric modeling and resulting dependencies that complete orientation information require is demanding, it is worth considering whether less specific proximity information could be substituted effectively. Arguably a person can recognize an image with distorted orientations among components, much as the subject of a cubist painting can be recognized.

Thus, if as I have argued, the network cannot be formulated with independence required by a tree structured decomposition, it is likely that local relationships – exclusion, co-incidence, proximity and orientation – can be exploited that do not require "completely global" influences.

IV. Single verses multiply connected networks

When more than one feature predict one lower level observable, the recognition network will have nodes with multiple paths. Fortunately the hierarchical method creates networks with influences only between adjacent levels. This section presents a solution to a simple case of multiple connections between levels that suggests their may be efficient solution techniques that apply to hierarchical networks.

A. The Concept of a solution

Both influence diagram and belief network solution methods result in the same solution to a Bayes network. Solving the network by influence diagram techniques derived from Bayes rule transforms the network so that a subset of the nodes, the set of hypotheses is conditioned upon the rest. The initial distribution over the set of hypotheses, known as the prior, is part of the network specification. In the process of solution, evidentiary nodes are transformed so that they are not conditioned by other nodes. Solving the network also imputes a distribution over the evidence, known as the *pre-posteriors* or *marginals*. The *pre-posteriors* – the distribution imputed by prior beliefs before observations are made – has significance for the solution only as it relates to collection of information – only indirect significance for the solution. As evidence nodes are instantiated,

their distributions are replaced by degenerate distributions (e.g. observed values), and repeated application of "Jeffrey's rule" changes the hypotheses to distributions posterior to the evidence. These posteriors are the results, by which choices can be made.

When solving by belief network operations, marginal distributions, or "beliefs" are maintained at all times for all nodes. Initially these are, for the hypotheses, the priors, and for the evidence, the pre-posteriors. As nodes are instantiated, message passing schemes update all node marginals to their distributions given the evidence observed.

The subset of singly connected Bayes networks to which Pearl's solution technique applies has the particular and useful property of being modular in space and proportional in time to the network diameter. There are several extensions for more complicated nets, by conditioning over cut sets [Pearl, 1985] and by star decomposition [Pearl 1986]. At the other extreme of complexity, Shachter [1986] shows that any directed acyclic probability graph has a solution in finite number of steps.

B. Networks with multiple hypotheses

When the various objects that may appear can each be decomposed into a different tree then these trees can be linked to a common set of leaf nodes. The result is no longer a tree, although it does not contain multiple directed paths. One may conjecture that when the leaves are instantiated and are no longer probabilistic, the trees effectively separate into a forest, and the solution is equivalent to evaluating each tree independently. Then Pearl's algorithm could be applied to each tree. This conjecture is wrong. As Pearl recognizes, [Pearl 1985] nodes with multiple parents cannot be the separating nodes in a cut-set. The parents of each leaf are not conditionally independent given the leaf node. This is apparent by application of Bayes rule through influence diagram transformations to condition the parent nodes upon the leaf node. Reversing a parent-to-leaf arc generates an arc between parents. This dependence is mediated in Pearl's algorithm by message passing at the leaf nodes.

In terms of message passing, incoming πs are reflected at leaves and affect the upward propagating λs even when the leaf value is certain, as shown by the propagation formula at an instantiated leaf node for λs:

(1) $$\lambda_i = \alpha \Sigma_j \pi_j p\{E_k / P_i P_j\},$$

[Pearl 86, from equation 21] where P_i is the parent receiving the lambda message, P_j is the parent sending the π message and k indicates the state at which the evidence is instantiated.

For separation such that multiple parents remain independent after evidence arrives, the evidence must be distributed thus:

(2) $$p\{E \mid P_1, P_2\} \propto p\{E \mid P_1\} p\{E \mid P_2\}.$$

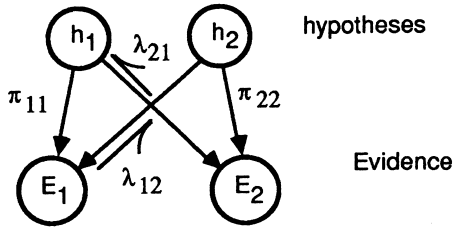
This is equivalent to evaluating each tree separately in trees where the leaves contain probabilities $p\{E \mid P_i\}$. Equivalently, this assumption makes the outgoing λs in (1) equal to the certainty value of the leaf node.

In exchange for the complexity that common leaf nodes add, we gain the ability to express exclusion and co-incidence among hypotheses of the different trees. This is the significance of (1): the $p\{E \mid P_1, P_2\}$ express the effect of one hypothesis on another.

C. Solving a simple multiply connected network.

Its a reasonable conjecture that some multiply connected networks lend themselves to time-efficient solution techniques. We offer one example where a multiply connected network can be solved by an extension of Pearl's technique. For this simple case consider a pair of common-leaved

trees only one level deep, with two hypotheses and two evidentiary nodes, shown this diagram:



This network has a closed form solution in terms of Pearl's messages. From the top-down propagation rule for root nodes [Pearl 1986, p. 260] the hypothesis sends down a π_{HE} message vector that is the term by term product of its prior with the λ message it receives:

(3) $$\pi_{HE} = \alpha \pi_i \lambda_i.$$

For each hypothesis node, write this as the product of a vector (all λs and πs now represent vectors) and a diagonal matrix $\underline{h}$ of the prior

(4) $$\pi_{22} = \alpha \underline{h}_2 \lambda_{12}$$

(5) $$\pi_{11} = \alpha \underline{h}_1 \lambda_{21}$$

where α normalizes π, and $\underline{h}$ is the matrix with the terms of h on its diagonal.

From (1) reflection equations at the leaves are, in vector notation:

(6) $$\lambda_{21} = \alpha P_2 \pi_{22}$$

(7) $$\lambda_{12} = \alpha P_1 \pi_{11}$$

where, as before, the likelihood, $P_i = p(E_{i_k} | h_1, h_2)$, is instantiated as indicated by k. So for each leaf node, the λ propagating up is the matrix product of the likelihood with the π it receives.

Substituting together (4)(5)(6)(7) obtains

(8) $$\pi_{11} = \alpha \underline{h}_1 P_2 \underline{h}_2 P_1 \pi_{11}.$$

This expresses the cycle of messages passed in one direction. For the other cycle containing π_{21}, the P matrix transpose appears in the corresponding leaf reflection equations.

Viewed as a recursion equation, if this has a solution, messages converge to steady values. Dividing through by the normalization constant puts this in the form of an eigenvector problem for the matrix $A = \underline{h}_1 P_2 \underline{h}_2 P_1$ with eigenvalue $1/\alpha$. Its eigenvector is the solution to the recursion relation. From this eigenvector, the beliefs of all nodes can be calculated. For example, let the evidence at both leaves have the same P matrices, and both $\underline{h}$ functions represent a uniform prior. Then

$$P = \begin{bmatrix} .52 & .18 \\ .08 & .22 \end{bmatrix} \quad Bel(h_1) = \begin{matrix} .811 \\ .190 \end{matrix}$$

$$\alpha = .562 \quad Bel(h_2) = \begin{matrix} .345 \\ .655 \end{matrix}$$

$$\pi = \begin{matrix} .811 \\ .190 \end{matrix}$$

Further simplifications occur for a rigid hierarchy. Notice that in a singly connected network we need only propagate upward to all nodes since determining the resulting marginal beliefs of evidentiary nodes is uninteresting as far as hypotheses distribution updates are concerned. Generally, with multiple connections above, we can ignore downward propagation in singly connected extents below. This follows from the lambda updating formula. As its shows, λs in a single parent node depend only on the λs arriving from below, and not on the node's marginal belief, nor on πs descending the tree. This is not symmetric with propagation downward; the πs interact with the λs on their way down. Thus, in a true tree, the hypothesis posterior can be updated solely by lambda propagation upward. This is equivalent to so-called "naive Bayes" updating schemes.

If we think of a network equivalent to the rigid hierarchy, but instead with all nodes at a level coalesced into one node, then the same argument about λ propagation applies: To update the hypotheses we need only propagate up between levels.

V. Further directions.

Empirical tests will determine whether the directions described in this paper improve the ability of vision based modeling systems. There are also a host of formulation and solution questions to pursue.

What is the relation of the attachment graph – the graph of geometric relations – and the graph that predicts observable features –the "recognition network" discussed here? They are related, but they are not the same thing. We have given arguments that the recognition network may have structural regularities that simplify its solution. The design of a general purpose vision machine also requires that the recognition network can be constructed in the process of recognition. The structure we have proposed has implications for the automatic generation of these networks, which we have yet to explore.

As we consider scenes with a wider variety of objects and the vision machine gains flexibility by having more models from which to chose, the recognition network becomes bushier. With even a small number of models it may not make sense to solve the whole network. There is a need for partial evaluation methods for large networks. One hopes also that the value and decision node structure of influence diagrams suggest techniques to control inference and to guide automatic generation of hypotheses over sets of evidence.

References.

[Binford 71]

Binford, Thomas O., "Visual Perception by Computer," **IEEE Conf. on Systems and Control**, Miami, (December 1971).

[Binford et al. 87]

Binford, Thomas O., Tod S. Levitt and Wallace B. Mann, "Bayesian Inference in Model-Based Machine Vision," in L.N. Kanal , Tod S. Levitt and John F. Lemmer, **Uncertainty in Artificial Intelligence 3,** (Elsevier Science Publishers, Amsterdam, 1989).

[Levitt 85]

Levitt, Tod S. "Probabilistic Conflict Resolution in Hierarchical Hypothesis Spaces," in L.N. Kanal and John F. Lemmer, **Uncertainty in Artificial Intelligence ,** (Elsevier Science Publishers, Amsterdam, 1986).

[Pearl 85]

Pearl, Judea,"A constraint propagation approach to probabilistic reasoning,"in L.N. Kanal and John F. Lemmer, **Uncertainty in Artificial Intelligence ,** (Elsevier Science Publishers, Amsterdam, 1986).

[Pearl 86]

Pearl, Judea, "Fusion, Propagation, and Structuring in Belief Networks" **Artificial Intelligence, 29** (1986) 241-288.

[Shachter 86]

Shachter, Ross, "Evaluating Influence Diagrams," **Operations Research, 34** (November - December 1986) 871-882.