Views on Vision

Eugene C. Freuder

Vision Group
A. I. Laboratory
M.I.T.

Introduction

In our discussion of the "object partition problem" we introduced a number of concepts which may be expanded upon to provide a unified theoretical base for a wider segment of our work in vision research.

This paper attempts to indicate some of the scope and subtlety of the issues involved. We discuss semantic, syntactic and contextual distinctions that bear on visual analysis. We focus on line predicate analysis, and indicate how various other aspects of the information content of the picture graph can be treated with the same structural approach we applied to partition analysis. This consistency is used to facilitate a view of scene analysis as an interlocking heterarchy of knowledge structures.

A very general discussion in Part I of issues which shape our approach, is followed in Part II by a solidification and expansion of the discussion in the context of some specific line predicates and the subproblems of visual analysis which they characterize.

(Familiarity with Vision Flash 4, The Object Partition Problem, is assumed in this paper.)

I.

Model:

Semantic/syntactic analysis is gaining currency as a basic conceptual tool in artificial intelligence. However care must

be taken if the application is to yield a net gain in insight. We will employ the well defined approach of mathematical logicians and model theorists to such distinctions. We will not develop this approach in great detail here, but rather, informally, in an attempt to clarify our view of the scene analysis problem.

We start with a syntactic "language" and a semantic "structure". The language will consist of the elements that form "line drawings" or "picture graphs", points, line segments, regions, plus predicate symbols to be discussed later. The structure will consist of physical corners, edges, faces, and various physical predicates. Our interest lies largely in the "interpretation" that maps constructs in the language, individual picture graphs, into physical constructs, three dimensional physical scenes.

We consider first individual picture graphs without attendant predicate symbols. We seek the physical scene that corresponds to the picture graph. We may decide that no physical interpretation is possible. More likely several interpretations are possible. The process of choosing the "best" interpretation, and obtaining further specific information about the nature of the corresponding physical configuration may be viewed as a problem in determining how appropriately to apply various predicate symbols to the picture graph.

An individual line, for example, may be interpreted as representing one or several physical edges. A predicate symbol

"$" might be applied to the line with the understanding that "$" is interpreted as the physical predicate "is a single edge". A priori this sort of label manufacturing may not seem to mean much. However, we have seen, in the partition problem paper, how convenient it may prove to work directly in picture terms, and we generalize this approach below. More importantly the predicate concept does not let us forget the semantic content of our syntactic manipulations. In particular the predicate approach stimulates the search for basic relationships that may be used to characterize or inform the more general questions of scene analysis. We will focus attention in this paper on a class of predicates we call "line predicates".

(In this paper we will take the somewhat inelegant approach of using these distinctions where we need them for clarity and blurring them where we need to for simplicity. This approach is facilitated by the observation that a physical scene, in a given view, can indeed be treated largely isomorphically with the corresponding two dimensional projection. In particular, we will deal largely with "line" predicate rather than "edge" predicate terminology.

The concept of scene plus view could be expanded on. We think it may be important to recognize that our low level three dimensional predicates need not concern themselves with "hidden" views concerning invariant relations, but may rather describe view dependent relations. The equating of semantic physical models with full three dimensional encodings may have needlessly hindered and confused the required fusion of syntax and

semantics in vision research work. At any rate it will be understood that we may still use the familiar terminology, "the line at the concave junction of the two regions", for example, as an abbreviation of "the edge represented by the line at the, etc.".)

Line Predicates:

A line predicate is simply a predicate that makes an assertion involving a line. It may seem tautological that we use lines as the basic unit of our analysis of "line drawings." However, attention has more often been focused on either the vertices or the regions of the picture graph. Our predicates will generally focus on the key properties that describe the relation of regions across line boundaries, as established at vertex junctions. In the case of the partition predicates, for example, the predicates concern the regions bordering a given line, but we generally are concerned with the application of these predicates to the set of lines entering a given vertex. Thus the various levels of picture graph elements are drawn together in the line predicate concept.

Some of the properties of lines that are of interest are:

definition or membership--whether they define or belong to one or both of the regions they bound, and if one, which;

junction--convex, concave, flat, or illusory;

multiplicity--whether they represent one physical edge or two.

When the edge is illusory we are interested in which plane is on top of which, and more generally when an edge defines only

one of the bounding regions we are concerned with which region is "over" or "occluding" the other.

There are relationships between these properties, some of them obvious. Illusory implies one edge; flat implies two edges; one edge implies that the associated regions belong to the same body, two edges the opposite.

Of equal importance perhaps, line predicates for these properties may be operated with in a uniform manner which facilitates the heterarchical interaction required of a successful vision system.

Analysis:

In brief, any line predicate may be given the treatment afforded the "partition predicates" as described in the previous paper. The immediate benefits may may vary from one predicate to another but the basic theoretical advantages remain, and increase as the aproach is extended to unify more of the vision system analysis.

We begin our treatment of a line predicate (or set of predicates) by establishing a "complete characterization" of possible interpretations of a picture with respect to the predicate property. From this base we can apply information from context and other knowledge structures to eliminate, dictate or rate on a plausibility scale the various choices we have in applying the predicate symbols to the various portions of a given picture graph. On this basis we can compile an analysis, or alternative analyses, of a given scene. We may arrive at an operational procedure that can be stated largely in

terms of syntactic rules for manipulating the predicate symbols. The procedure will derive, however, from a semantic motivation, as it seeks to provide a semantic (physical) analysis.

A complete characterization is initiated with an enumeration of all possible intepretations of the various vertex configurations with respect to a given predicate (or set of predicates). Syntactically we can list all possible labellings of the vertex type with the corresponding predicate symbols. However, we examine the semantic predicate concept for posible restrictions; the partition predicate was observed to be an equivalence relation, for example, and so we established a corresponding syntactic restriction. Symmetries within a vertex type will tend to reduce the number of different labelling possibilities as well.

The physical domain will very likely be restricted in some fashion, e.g. to planar polyhedral solids. This may impose further restrictions on the predicates. We will not be concerned in this paper with any context more general than planar polyhedra, i.e. we are not concerned with curved objects. In this context we can observe that any line (segment) must be consistently labelled along its length for any of the predicates we consider. The line cannot receive one interpretation at one endpoint, another at the opposite endpoint; e.g. a line cannot change from a concave to a convex junction.

Our "complete characterization" should be undertaken at the most general possible context level to begin with. In our case we suggest a physical domain limited to planar polhedra, with

few initial constraints on the picture graph or interpretation.
Our most general context should allow the various "degenerate"
occlusions and overlaps of lines and vertices. Certain of our
specialists or initial programs may not be able to handle these
problems, but they should certainly be included in the overall
universe which our system must ultimately understand.
Establishing an optimal level for the most general context under
a given approach is itself a research decision. General views
of planar polyhedra seems historically and technically to be a
good top level context for line drawing analysis.

At this level, after the semantic, symmetric and contextual
simplifications have been made, we are left with a list of
alternative vertex labellings for any given set of line
predicates and range of vertex types. A complete set of
alternative interpretations of any given picture graph may then
be obtained from applying the various vertex labellings to the
vertices of the picture graph in all consistent combinations.
That is we must observe the "one line, one label" criteria, we
cannot accept an overall labelling that requires any individual
vertices to receive illegal labellings, and we must avoid
"global inconsistencies" as in the case of partition labelling.

This achieves a complete characterization of a given
property or set of properties in picture graph analysis. Often
this much may buy us more than we expect. The possibilities may
be considerably narrowed down, basic questions such as "is there
any possible interpretation" will have been answered. At the
least we have a basis for organizing and directing our thinking,

and that of our programs.

We are, of course, generally interested in the question of which is the "best" or "most likely" interpretation of a given picture graph. We need to build on our characterization to provide an answer to that need.

Context:

Contextual issues arise in several forms in the restriction or ordering of the decision space in a given line predicate analysis. Often we are interested in what the most likely interpretation will be (or if any interpretation is possible) under a restricted context.

Further context restrictions may further narrow down or order the predicate labelling possibilities. A predicate analyser may thus be informed by, or alternatively inform, the context analyser of a heterarchical system. The ultimate goal, of course, is a system which can handle any context, employing "specialists" perhaps, on the advice of the context analyser. In the course of development of such a system, we often need to restrict the context of analysis, either to produce a specialist, or merely to simplify a problem sufficiently to get a foothold on it. It is imperative at such times that we fully appreciate the nature and implications of the context restrictions we make. To place this problem in focus we return briefly to our model theoretic viewpoint.

Restrictions may be imposed on either the syntactic or semantic domain, or on the interpretation. We may stipulate, for example, that only picture vertices joining three or fewer

lines may occur in a picture graph; that only trihedral corners may appear in the physical domain; or that three line picture vertices are to be interpreted as trihedral corners. These three possibilities are not identical. If we are to appreciate the difficulty of our problems and the extent of our accomplishments we must understand the nature of our restrictions. Huffman's "general position" criteria, for example, while a nice concept, is misleading in that it does not embody all of the restrictions Huffman apperently expects it to. A common problem of newcomers to the area of partition analysis is the imposition of restrictions that effectively reduce the problem to the trivial case where separate objects meet only at T-joint occlusions. This reduction need not be an explicit restriction; it follows from quite natural lower level restrictions. Furthermore, exclusive T-joint demarcation is a quite common physical situation. It happens, however, that it is almost a degenerate case as regards the full problem of partition analysis. Unless one realizes this, there is the temptation to possibly misuse the principle of generalization in concluding that the methods used in the restricted case extend or have analogues in the general case. The "simple" T-joint, for example, has possibilities in a more general context that make it impossible to use in any straightforward fashion as an indication of partition; on the contrary it may prove to be one of the most misleading and complex vertices to analyze.

Furthermore, we might note that the nature of the context restrictions we make, in order to achieve various results, may

at times be among the most interesting aspects of our analysis. As in mathematics, the required strengths, alternatives, and other considerations involved in various hypothesis restrictions, can constitute a significant aspect of a result. An appreciation of the context restrictions may be required, moreover, for an adequate theoretical model or practical realization of a comprehensive analysis, a model that can deal as effectively as possible with both restricted and unrestricted contexts.

Insofar as they affect possible or likely labelling choices for a given predicate, context restrictions can be viewed in at least two alternative fashions. They may formally define the area in which we are to ask our questions, for example the question of whether any physical realization is possible in a given context. They may heuristically define our decision space in order to limit or order it, with the option of lifting or changing such assumptive restrictions should they lead to unsatisfactory results. In the latter case, the choice of context restrictions, based on "general experience", specific contextual analysis of a given scene, or whatever, becomes a vital facility.

In summary, our context restrictions help to direct and define our results, and we cannot properly assess or extend these results without a clear analysis of these contextual assumptions.

Semantics:

Within any context, whether it be the most general context

or some more restricted one, we use semantic information. The nature of the semantic structure indicates the limits on the set of possible labelling choices in constructing a characterization of a given predicate. Semantic information also assists our judgment on the relative plausibility of the various choices, in order to inform our "best possible interpretation" decision procedure.

There seems to be some needless confusion on the semantic/syntactic distinction at this point. Again we may benefit from a formal analysis. In logic, a "theory" is a collection of statements, in a syntactic language, which may be defined either semantically, as statements which are true in some semantic domain, or syntactically, as statements derivable from specified syntactic "axioms" by specified syntactic rules. Often the interesting "metatheoretical" problem arises, whether a semantically defined theory can be derived also syntactically . The axioms and rules that are employed in a syntactic formulation of a semantically defined theory are not, of course, pulled out of a random hat. They are chosen with a keen awareness of the semantic content they will embody and must imply.

Similarly in our work in artificial intelligence. We may seek a "syntactic" procedure for simulating a semantic function, but there should be no surprise that our syntactic operations must be semantically informed to be reasonable. In natural language study, for example, the legitimate sentences of a natural language may be regarded as a semantically defined

theory. Some aspects of this theory are particularly amenable to syntactic formalization; these we generally term "grammar" or "syntax". This use of the term "syntax" may be misleading, and we are further hindered by a correspondingly restricted definition of language "semantics". However, in our more precise model theoretic semantic/syntactic terminology, there need be no surprise that a reasonable syntactic "language grammar" must implicitly, or preferably explicitly, acknowledge the semantic motivation for the theory it is attempting to generate. And there need be no surprise if a successful semantically informed approach to language analysis provides a simultaneous syntactic basis for "language grammar" and "language semantics."

Syntactic formulations sometimes generate their own momentum. They may indeed suggest new semantic insights. However, we must keep semantic motivations in mind if we are to develop a comprehensive and comprehensible syntactic model.

Knowledge Structures:

Semantic context implications are certainly important in determining plausibilities in our scene analysis. Knowledge of many sorts, however, may have a bearing on our decisions for any particular predicate or set of predicates. We face twin problems. We require the individual identification and analysis of these various aspects of visual knowledge of a scene. And we need to employ the heterarchical interrelation of the knowledge structures, in their construction, for a given picture graph analysis. We may be aided in our implementation of the latter

step if our individual analyses exhibit some common form.

The partition analysis paper discussed in some detail the advantages of a line predicate, complete characterization approach to the description and analysis of a given knowledge structure. We noted, in particular, how the labelling network description built in the potentially global interrelationship of the picture parts, and the complete characterization allowed an alternative failure flexibility. The predicate labelling aproach allows us to use any and all parts of the network for information, and both positive and negative indications.

We are interested in this paper in the multiplicity of knowledge structures which may be approached with this type of analysis, the observation that there are several line predicates, or sets of predicates, which are worthy of and amenable to an analysis of this sort. We believe that much of the information content of picture graphs can be profitably characterized thus by line predicates.

Certain sets of these predicates may be useful as basic units of analysis for various specific subproblems in the area of visual analysis, e.g. the partition problem. There are several qualities which might recommend a set of predicates as basic analytical concepts for dealing with a particular problem. We would like the predicates to characterize the problem in some sense. Ideally we would like the predicates to "fully" characterize the problem, in the sense that all possible interpretations of a given picture graph, with respect to the given problem, will correspond to alternative "labellings" of

the picture graph with the specified predicate symbols, and vice versa. A full 1-1 characterization in this sense allows the complete characterization of the predicate set to constitute a complete characterization of the specified problem, in the sense of all possible interpretations of the picture graph with respect to that problem.

The usual advantages of line predicate network analysis will carry over to a predicate characterized problem. In particular the predicate network is a convenient framework on which to add the context and knowledge information that we need to make our analysis choices.

If we have chosen our characterizing predicates well, they will indeed seem to reflect the fundamental properties of the problem under consideration. This can be most revealing as to the basic nature of the problem, and other relevant but secondary or separate considerations may then be brought to bear in a natural and effective model. Isolation and observation of the basic criteria of a visual property, and observation of their modification under context restrictions, and interdependent properties, is an ideal outline for understanding, and a procedural model which follows naturally from our basic theoretical approach.

If indeed many of our knowledge structures have the same predicate network labelling structure, the crucial process of conjoining the various structures heterarchically may be made that much more tractable. Our model of our procedural implementation of the visual process will be clarified.

II.

Without descending too far from the (ahem) high level of generality intended for this paper we might indicate some of the areas in which we are studying line predicates other than partition analysis, in order to illustrate and elaborate some of the issues and concepts discussed here.

Junction:

Huffman has done some work with what we would term the line predicates concavity, convexity and "illusory." (Huffman actually uses two predicates which combine the illusory concept with an indication of the true definition of the edge.) These predicates have great relevance to the partition problem and the "possible configuration" or "existence" problem--whether a given picture graph has any realizable interpretation as a physical scene. Their usefulness will be clearer with the inclusion of the "flat" predicate (junction of two coplanar planes), and a complete characterization approach (Huffman begins with a strongly restricted context). However, we believe that the predicates concave, convex, flat, illusory, characterize what might be termed the "junction problem", and are not appropriate as basic units of analysis for the partition or possible configuration problem. They do not characterize, or fully characterize those problems in the manner that we indicated we would like a set of predicates to characterize a visual analysis problem.

It should be obvious, from our partition problem paper, that any attempt to use the junction predicates as basic units of partition analysis is unwarranted. A complete characterization could not even be attempted.

The issues involved in the characterization of the junction problem could be expanded upon at some length. However, it might be more instructive for the purposes of this paper to choose the "possible configuration problem" for further discussion.

Configuration:

We note first that the "possible configuration" or "existence" problem is in a sense an aspect of a general "configuration" problem. In the case of partition analysis, recall, we were interested first in characterizing all possible (physically realizable) partitions, then in choosing the best one. For the configuration problem the interest in characterizing all possible configurations may lie largely in determining if any configuration is possible, i.e. the existence problem. The question of choosing the "best" configuration is, of course, of interest. However, since this question essentially asks for a complete analysis of the scene, its answer is generally the result of previous decisions on the various aspects of the scene analysis.

This interpretation of the "configuration" problem will become more concrete if we choose an appropriate formalization of the configuration problem, e.g. in terms of line predicates.

Huffman is, of course, not dealing in our characterization

framework, and he clearly presents his junction labelling approach as one among several designed to bear upon the existence problem. We have no argument there, of course. We wish to consider whether the junction predicates would be appropriate for our characterization approach to the configuration problem, or whether another predicate set should be proposed as a basic analytical characterization, on which such relevant advice as provided by the junction knowledge structure might be hung.

Huffman indicates that the junction predicates are not used as what we would term a complete characterization of this problem. In any case, they do not appear suitable to assume even something less than a "full" characterization role with respect to the configuration problem. A proper choice of characterizing predicates for a given problem will, as we have indicated, hopefully address themselves directly to the basic nature of the problem. The junction predicates do not meet this criteria. An attempt to view them as characterizing, rather than relating to, the problem will indicate that there are more basic predicates, assumed in the definition of convex, etc., that really seem to be the criteria of a possible physical realization.

Context:

the basic issues involved should become clearer if we consider the configuration problem its most general context. In beginning with the restricted context Huffman deals in, we lose or obscure some of these issues.

Guzman distinguishes pictures that have no possible physical interpretation as those which could not be produced by photographing or projecting any three dimensional scene from any view. This definition seems desirable as a basic characterization of an "impossible" object.

Guzman notes that many of the standard "optical illusion" "impossible" figures do not fall under this definition, however. The "Penrose Triangle", for example, can be obtained as a special two dimensional projection of at least two different three dimensional constructs. The study of such "optical illusions", or "impossible figures", precisely formulated, becomes then a question of which picture graphs have a physical interpretation, given certain contextual constraints, on the physical domain, the picture domain or the interpretation. This is the problem that Huffman sets himself, for a context which he defines.

Under these circumstances the nature of the context restrictions one can consider becomes one of the most interesting aspects of the study. In a sense they form the content of a reasonable model of what constitutes the "illusion".

Characterization:

We now have an idea of the general nature of the configuration problem to which we wish to apply a line predicate characterization. We begin by seeking a basic characterizing predicate for the configuration problem, starting at the general context level of arbitrary views of planar polyhedra. Context

restrictions may then be adduced in an instructive fashion. We might tentatively propose the following line predicate: l belongs to the boundary of R1 as defined by R2, or briefly R2 defines R1 along l, where R1, R2 are faces, bordering an edge l in a given view of the scene, and l physically forms a bounding edge of R1. We might employ a predicate symbol "→", for example, to be interpreted as this predicate, which we refer to as the "definition" predicate. The arrow would be drawn at right angles to a picture line, which could be labelled with a single arrow in either direction or with two arrows in opposite directions (abbreviated as a double ended arrow). Consider the following labelled figure:
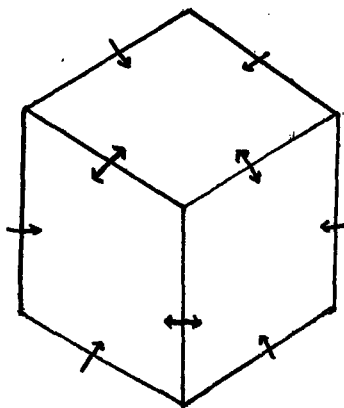
We do not intend to present a complete analysis of this predicate here; indeed we have not yet fully satisfied ourselves it is the ideal predicate for our purposes. However, some further discussion should be instructive.

From our most general semantic context we observe that this "definition" predicate is anti-reflexive; R0→R0 is impossible. Correspondingly any label for a line which is bordered on both sides by the same region is impossible. This criteria will, we believe, encompass all the truly "impossible" objects in our

general context, and allow us a complete characterization of the possible configuration problem.

Restricted Context:

As it happens the context considerations here are very much the opposite of that in the partition problem, for example, in the following sense. For the configuration problem we enrich the problem, and discover its most interesting and difficult aspects, not by allowing the most general context, but by considering a restricted context. We may justify enriching the problem in this fashion by appealing to an interest in the "human" optical illusions, or by noting the advantageous aspects of employing and understanding the implications of such context restrictions.

In particular the context restrictions or assumptions that lead humans to perceive an optical illusion (of the "impossible object" or other type)--right angle and straight line assumptions, absence of colinear overlapping lines, whatever-- should not be dismissed as weaknesses in the human perceptual system. More likely they are very good heuristics that facilitate human processing and description of visual experience. They constitute plausibility rankings for the interpretation decision structure, based in part on a knowledge of context plausibilities derived from wide visual experience in a particular environment.

In summary, strong context plausibility judgments, while very useful in choosing a "best" interpretation, may essentially function as at least a temporary context restriction, which may

lead to "optical illusions", or "impossible figures". Thus our interest in determining useful context based plausibility criteria may well benefit from a study of context restricted "impossible figures". (Though more interest should therefore be shown, than has been at times, in the various alternative context restrictions themselves, which can be made to "model" various human "optical illusions".)

Consider the simple figure:
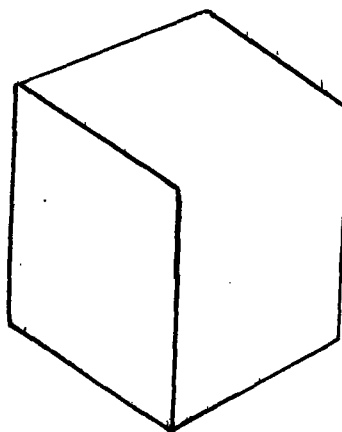
This is certainly no impossible object in any deep sense. We could easily be looking straight down at the top of two nested bricks, for example, one cube, one L-shaped. If we were looking at the bases of two pyramids we could shift our viewpoint a fair amount without changing this projection much. The interesting part here is in fact the contextual experience, and the corresponding context assumptions and context based judgments, that might lead us to choose as our most plausible interpretation a cube with a missing line, and should correspondingly lead a vision system to call in its line verifier. (This raises obvious similar issues for our consideration of the junction problem and the junction predicate.) A complete vision system would, of course, be aware

of the possibility of an "overhead view" to fall back on.
Similarly a knowledge of the restrictions that lead to the
Penrose Triangle illusion gives us, or a vision system, a good
idea of how to physically realize the Triangle in a more general
context.

Restricted Characterization:

Very briefly now, various alternative context restrictions
can be proposed which permit formulation of principles which
embody considerable restrictions on the number of possible
"definition predicate" labellings. This enables us to catch
certain "context impossible" configurations, e.g. as in the
above figure. We require rather sophisticated semantic
observations, however, to fully restrict the possible
labellings, and to assure "global consistency". We may also
employ other knowledge structures, such the junction predicate
labelling network.

Huffman presents several subtle criteria that can be viewed
in these roles. Optimally, if our characterizing predicate
choice was ideal, any criteria that implies a picture graph is
not physically realizable under the given context restrictions
will also imply that no predicate labelling is possible that is
consistent with the given context.

This issue may be clearer in the familiar context of
partition analysis. By applying certain syntactic restrictions,
derived from the general semantic context we set for ourselves,
to the partition predicate labelling procedure, we achieved a
"complete characterization" of the partition problem. That

meant that for the general context "structure" we had achieved a 1-1 correspondence between syntactically allowable labellings and semantically (physically) possible partitions. (Technically a "complete" and "consistent" characterization.)

We then observed that restricting the context further could indicate further restrictions on the physically possible partitions, and corresponding syntactic restrictions could be made on the labelling possibilities. These might take the form of general principles (such as the "transitive" property used in the general context) or simple elimination of certain labelling possibilities. We did not concern ourselves at that time with proving that for any or all given context restrictions a corresponding nice (e.g. finite) set of syntactic restrictions could be found that would insure there was still a 1-1 relationship between possible syntactic labellings and possible physical configurations. That is, there is the question of whether we can continue to have a complete characterization of the partition problem in various restricted contexts. Applied to the configuration problem this issue becomes more vital, since in this case, as we have noted, interest lies in the restricted characterizations, whereas in partition analysis the general context forms the basis of our problem.

Conclusion

Well clearly we are concluding this paper by beginning several others, as we concluded the partition paper by leading

into this discussion. But then, the continuing implications of this work are indeed part of the point of this paper. Our "definition" predicate study, for example, could be viewed, perhaps, partly as a formalization of some recent work of Patrick Winston, on physically associated edges in shape determination, as the partition predicate analysis formalizes the germinal work of Guzman and others in partition analysis.

Hopefully we have provided some idea of the scope and significance of a number of the key concepts, observations and methods that are guiding our current vision research.

Acknowledgements

The theoretical approach described in this paper is a result of a happy confluence of a propitious theoretical atmosphere and the working environment of the Vision Group Vision System.

Among those who contributed to either the atmosphere or the environment (Ralph Nader excepted):

M.   Dowson

A.   Griffith

A.   Guzman

B.   Horn

D.   Huffman

J.   Lerman

M.   Minsky

S.   Papert

M.   Rattner

P.   Winston