GENERAL PURPOSE MODELS: EXPECTATIONS ABOUT THE UNEXPECTED

Steven W. Zucker, Azriel Rosenfeld, and Larry S Davis
Computer Science Center
University of Maryland
College Park, MD 20742 U.S.A.

Abstract

How can computer vision systems be designed to handle unfamiliar or unexpected scenes? Many of the current systems cope quite well with limited visual worlds, by making use of specialized knowledge about these worlds- But if we want to expand these systems to handle a wide range of visual domains, it will not be enough to simply employ a large number of specialized models. It will also be important to make use of what might be called "general-purpose models." Such models can be used to suggest reasonable descriptions for a given scene, e.g., in terms of features or regions or groups of these, in the absence of specific context.

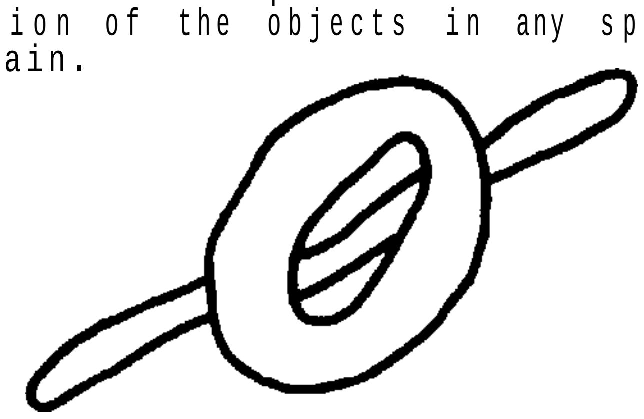
1, The Need for GPM's

In recent years, many computer vision systems have been constructed. These cope quite well with limited visual worlds such as the blocks world (Shirai [1]) or office scenes (Garvey and Tenenbaum [2]). They do this by making use of specialized knowledge about the worlds in which they operate. An important goal of future computer vision systems is to expand this capability to cover a wider range of visual domains. But can this be done by simply employing larger numbers of specialized models? The contention of this paper is that it will also be important to make use of what we have called "general-purpose models" (GPM's). By definition, these are models which are applicable even when we have little or no a priori knowledge about the class of scenes that is to be analyzed. They include models for general classes of local features (edges, lines, angles, etc.) that occur in many different types of scenes, as well as models that describe how such features can be grouped into aggregates. (These aggregates may in fact not correspond to objects but they can serve as useful first guesses to guide later steps in the analysis.)

Before describing GPM's in any detail, we can illustrate the need for them by a few examples. Imagine a computer vision system attempting to deal with scenes chosen randomly from a highly varied collection – pictures of rooms, buildings, faces, scenery, etc. It would seem a poor strategy for the vision system to start by asking specific questions such as "Is there a chair in the *scene?"* or "Is there a tree?". By analogy with the game of 20 questions, a more reasonable strategy would be to ask general questions which would aid in reducing the search for appropriate global frames (Minsky [3]). An example of a general question might be "Are there many straight edges in the scene?" This could serve as a first basis for discrimination between "natural" and "man-made" scenes. The knowledge implicit in this question is a reasonable generalization that can be made on the basis of experience with a variety of scenes, and illustrates one way in which GPM's can be used in scene analysis.

Because the types of pictures mentioned above are familiar to us, it is difficult to separate the effects of special purpose knowledge from those of general purpose knowledge. However, there are many types of pictures that are unfamiliar except to specialists -- electron micrographs, for example. Yet in spite of their unfamiliarity, if one asks people to sketch the significant regions in such a picture, the resulting sketches will usually be highly consistent. This ability to impose structure on unfamiliar scenes reflects another possible role of GPM's in scene analysis. The model, or class of models, implicit in this example can be motivated by observing that most scenes contain objects that differ physically from their surroundings, and this physical difference usually results in a visual difference. Thus, we find it reasonable, in sketching an unfamiliar scene, to attempt to divide it into regions that are more or less homogeneous in brightness, color, or texture.

The line drawing shown below illustrates our point at a somewhat different level. Here, the significant regions are already well defined, but GPM's can play a role in grouping these regions into rather schematic objects, and in perceiving the relations between these objects (e.g., behind, next to). This process *is* independent of the interpretation of the objects in any specifie domain.



2. The Nature of GPM's

It is widely accepted that any scene analysis system should utilize whatever specific knowledge is available about the class of scenes that it is to "understand" (see, e.g., Minsky and Papert [4]). One of the fundamental problems facing the designer of a scene analysis system is the representation and effective use of such knowledge. Several powerful paradigms have been offered to guide the designer in solving this problem – e.g., those of Winograd [5] (procedures) and Hewitt et al. [6] (actor formalisms). (See also Pylyshyn [7] for a review and critique of such representation schemes.) An important feature of these paradigms is the structuring of knowledge into models. For example, a model of a chair might specify that the chair has a seat, legs and a back, while a model for gravity could specify that any object must be supported. The conditions required by a model can be expressed as goals that a computer vision system can attempt to satisfy. Programming languages such as PLANNER [8] and CONNIVER [9] were specifically designed to accommodate such goal-directed processing.

Intuitively, there are many different types of goals. For example, the rather specific goal "Find a table leg - I think that the table top is at position (x,y,z)" is different from the more general goal "Find a straight edge." Corresponding to these different types of goals are different types of models. Specific or specialized models correspond to things like cubes or chairs. These are useful when a good deal is conjectured or known a priori, or has been found in the course of previous analysis (cf. Haber 110] for some important psychological parallels). GPM's, on the other hand, are models that are applicable even if little or nothing is known a priori, and analysis has not yet begun.

It should be stressed that there is no sharp dividing line between specialized models and GPM's, nor is there even a simple linear ordering of models with respect to their degree of generality. One might at most hope for a partial ordering: e.g., models for general classes of features or objects could be regarded as GPM's in comparison with models for more specialized classes. Nevertheless, we have used the term "GPM" in this paper for the sake of simplicity.

2.1 Feature Detectors

As already mentioned, an important class of GPM's consists of models for various types of local features – edges, lines, angles, etc. Such features occur in a wide variety of scenes, so that detectors for these features will be very generally applicable. Moreover, such features are often relevant to the description of the given scene, because they usually arise from objects that are present in the scene. Indeed, since objects are distinct from their environment, there should be a discontinuity of some sort where an object ends and the background begins. In particular, there is likely to be a change in brightness or reflectance, so that visual edges will be present. Edges can also serve as a basis for a simplified description of the scene, since regions in the scene can often be approximated by specifying their edges and the colors, or textures, of their interiors. (For evidence that the human visual system may encode regions in this way see Cornsweet [11].) In summary, models for edges are an important class of GPM's because (a) they are widely applicable; (b) they often serve to delineate physical objects; and (c) they provide a basis for giving a simplified description.

Similar remarks can be made about models for other types of local features. Angles, for example, represent abrupt changes in slope, just as edges represent abrupt changes in gray level, color, or texture. They are thus likely to be significant features of the shape of a region. At the same time, as pointed out many years ago by Attneave [12], angles provide a basis for giving simplified description of shapes, since the shapes can be approximated by polygons whose vertices are located at the angles. In the light of these remarks it is not surprising that detectors for local features such as edges and angles have been found in many biological systems [13].

2.2 Grouping Rules

GPM's should also encompass mechanisms that can deal with the abundant information obtained from feature detectors and put this information into a more useful form. Such mechanisms might, for example, serve to group features into regions, or to select salient features. Julesz [14] has proposed a clustering scheme in which proximal points having similar feature values group into clusters. Such a clustering process is similar to many of the region growing schemes proposed for scene analysis systems (e.g., Muerle [15]). Rosenfeld [16] has suggested a process of best feature size determination and non-maximum supression that can be used to select locally salient features such as edges, spots, and streaks.

The Gestalt psychologists, notably Koffka [17] and Wertheimer [18], provided demonstrations for the existence of GPM's of these general types in human vision, although they did not offer adequate mechanisms for these models. Their organizational laws of good continuation, similarity, closure, and proximity describe qualitatively how the human visual system prefers to group features into aggregates. The Julesz clustering process mentioned above can be considered a first step toward modeling similarity grouping. Beck [19] has conducted studies of the factors that determine similarity grouping and has related these factors to the concept of peripheral discriminability. These ideas can also be incorporated into even more comprehensive texture models (Zucker [20]).

The Gestalt notion of good continuation deals with intersecting curves. It can be interpreted as a preference for pairing off the curves at an intersection in a way which minimizes the total curvature. The importance of these more global, yet still rather unspecialized, mechanisms for organizing even just a brief "glance" at a scene has recently been demonstrated by Weisstein and Harris [21].

The Gestalt laws, like the feature detectors, can be rationalized as being well matched to certain aspects of physical reality. Objects are often physically homogeneous, and a similarity grouping is thus likely to arise from a single object or population of related objects, so that it is sensible to perceive the grouping as a unit. Physical objects are normally compact, so that a proximity grouping is likely to arise from a single object, and it is thus reasonable to perceive it as a unit. Discrete physical objects must have closed boundaries, so that the law of closure is also a sensible heuristic. Thus, models whose effects are similar to those of the Gestalt laws would be useful as GPM's, since they would be widely applicable, and would give rise to regions that often correspond to objects, or useful pieces of objects. At the same time these models would provide a basis for further simplifying the scene description, by determining aggregates of features that can serve as proposed constituents for attempting a more realistic semantic interpretation of the scene. In other words, these models serve to significantly reduce the combinatorics involved in determining the possible groupings of the features in a scene, by making only a few combinations easy to "see", while the other combinations are "hidden figures."

2.3 Depth Cues

As another class of examples of GPM's, we will briefly consider some of the cues, and the associated processing, which can be used to deduce depth relationships of objects in a scene. These GPM's are somewhat more specialized than those discussed previously, since they are predicated upon having delimited objects of some sort. Our discussion of the simplifying nature of Gestalt laws and depth cues in Sections 2.2-3 is based in part on Hochberg [22],

The visual field is two-dimensional; many different spatial arrangements of objects can give rise to identical visual images. Nevertheless, under normal circumstances one has little difficulty in judging the three-dimensional positions of objects, even from a single monocular visual image. In making such judgments, one uses a variety of cues which help to structure the visual field in an unambiguous way with respect to depth. Familiarity is certainly important in determining depth relationships — if we can recognize an object, then we know its actual size, and can calculate its depth from its apparent size. Many depth cues, however, do not depend on familiarity; rather, they reflect generalized knowledge about objects. If two similarly shaped objects have different apparent sizes, we tend to judge the smaller one as farther away. This amounts to assuming that the two objects are actually identical, and are therefore the same size. This first order default option is useful in the absence of any other information. Note also that this assumption leads us to a "simpler" description of the scene as containing two identical objects at different distances, rather than two dissimilar objects. Similar remarks apply if we have two regions containing the same type of texture, but with one more coarsely textured than the other. In this case the simplifying assumption is that both objects have identical textures, but one is farther away.

More generally, if an object is asymmetrical, but can be regarded as a three-dimensional rotation of a symmetrical object, we tend to judge it that way; for example, we can interpret a trapezoid as a rectangle seen obliquely. This perspective interpretation yields a simpler description of the object, namely symmetrical but rotated, rather than asymmetrical. At the same time, the perspective assumption is a reasonable one to make, since it would be an improbable coincidence for an object to appear symmetrical when rotated, if the object were not actually symmetrical. Analogously, if we have a region containing a texture whose coarseness varies from one side to the other, it is easy to see the region as uniformly textured, but tilted. This gives the scene a simpler description, consistent with our knowledge of the physical world.

As a final example, suppose that an object has an irregular shape, but that it could be symmetrical, or congruent to some other object, if we regarded it as partially hidden by a neighboring object; then we will tend to conclude that this is actually the case. This interpretation in terms of interposition yields a simpler description, and it certainly represents a situation that frequently arises in practice. Interpositions are often (but not always) signalled by local cues such as "T-junctions": the object whose edge lies along the leg of the T disappears behind the object whose edge lies on the horizontal bar of the T. The disappearance and reappearance of one object from behind another may be indicated by the presence of "matching T's": T-junctions whose legs are good continuations of each other. It is clear how these cues generalize our experience with real objects.

2.4 GPM's in Human Vision

Analogs of GPM's can be found in many models of visual perception developed by psychologists. Bruner [23] theorized that early stages of clue extraction and object isolation direct later stages of processing. Preliminary operations of this sort on the stimulus "that produce the objects which later mechanisms will fill out and interpret" have been called pre-attentive processes by Neisser [24]. An important function of these processes is to "form the units to which perception may then be directed... Visual objects are identified only after they have been segmented, one from the other. This permits the perceiver to allot most of his cognitive resources to a suitably chosen part of the field... Since the processes of focal attention cannot operate on the whole visual field simultaneously, they can come into play only after preliminary operations have first segregated the figural units involved. These preliminary operations... correspond in part to what the Gestalt psychologists called 'autochthonous forces'..." (Neisser [24], pp. 86, 88, 89). Of course, it should be stressed that the "objects" extracted by the preliminary operations may not be the same as the final objects obtained by subsequent analysis.

The belief is certainly widespread among psychologists that the early stages of visual information processing do not depend on one's knowledge or expectations ("perceptual set") about the particular situation. "The effects of perceptual learning consist of changes in where you look, and of how you remember what you saw, but not of what you see in any momentary glance... Most of the attempts to demonstrate the effects of motivation and of set on perception... were memorial rather than perceptual" (Hochberg [25], p. 326). Thus, some of the operations performed by the human visual system can be thought of as corresponding to the use of GPM's at the early stages of scene analysis.

In thinking about GPM's in visual perception it is sometimes useful to imagine that they evolved as generalizations of more specific models for concrete physical situations by a process which gradually relaxed the domain-dependent context, so that semantic notions gradually acquired a syntactic flavor. In other words, semantics became "fossilized" and turned into syntax.

3. The Role of GPM's

As the universe in which computer vision systems operate becomes richer, the initial expectations of the systems will have to become less specific. Thus these systems will have to rely more heavily on GPM's, especially during the initial stages of scene analysis. "Thinking always begins with suggestive but imperfect plans and images; these are progressively replaced by better -- but usually still imperfect — ideas" (Minsky [3]).

To further illustrate the use of GPM's, as well as their interaction with more specialized models, consider Shirai's [1] scene analysis system. This system operates in the blocks world, and presupposes that any block will sharply contrast with the background. This knowledge is embodied in a GPM of object-surround contrast which performs a Kelly-like [26] edge detection of the outer boundary of the set of blocks. Next, it is assumed that all curvature maxima on the outer border of the set of blocks indicate vertices of blocks. Each hypothesized vertex is examined by procedures that contain specialized knowledge about the forms of vertices in the blocks world. These specialists can, in turn, suggest internal lines and vertices. These suggestions invoke GPM's for angles and lines to discover the hypothesized, but as yet unverified, internal features. The dynamics of this system's operation illustrate clearly the interplay between specialized models and GPM's.

Shirai's use of GPM's is based on the sharp contrast of the blocks with the surround and the fact that all angles correspond to salient features in the blocks world. For example, early in the processing an "object" interpretable as the outer contour of the collection of blocks is found. In general, scenes are not this simple, and the notion of what constitutes "saliency" is rather elusive. How might specialized models and GPM's interact in more complicated situations?

The more specialized the knowledge that we have about the scene to be analyzed, the more efficiently the analysis can be performed using specialized models. In an extreme case, if we know what objects are supposed to be present and in what positions, the analysis reduces to a simple verification of these facts, perhaps by template matching.

This is an example of a situation where GPM's would be unnecessary. More commonly, however, we may know the allowable types of objects, but not their positions, as in the blocks-world situation – or, we may know only that some sort of objects may be present, but not what shapes (or textures) these objects will have, as in the random slide show situation of Section 1. As hae been already mentioned, one could, in principle, run through many (thousands?) of specialized object models, applying each one to the scene in turn. A more efficient strategy would presumably be to initially apply highly general models, where the chance of success is greater.

By definition, GPM's should be useful in a relatively wide variety of cases. In fact, it may turn out that the most efficient way to start the analysis of a scene about which specialized knowledge *is* available is still on a general level. For example, in analyzing scenes containing polyhedra, it is more efficient to begin by looking for edges, as Shirai did, rather than for long straight edges or corners. This is because it is computationally more costly to detect corners (etc.) than edges, and these detectors will have to be applied over large regions since we do not know in advance where the polyhedra are. Once edges have been found, the more specialized knowledge that is available can be brought to bear by checking the edges for straightness. Thus, GPM's, because of their relative simplicity, are likely to be useful even when we do have special knowledge about the situation.

It should be apparent from this discussion that the use of GPM's does not presuppose any particular type of control structure in the given scene analysis system. A system that employs GPM's need not be "bottom-up" or hierarchically organized. GPM's are equally compatible with heterarchically structured systems, such as Shirai's. Furthermore, they are not necessarily the lowest-level goals in such a system, but may occur throughout.

4. Concluding Remarks

Line drawings such as that shown at the end of Section 1 constitute a possible domain for the study of GPM's. Here the GPM's would structure the drawings into (unfamiliar) objects having a limited set of properties (e.g., opaque/transparent, blob-like/elongated) and satisfying a limited number of relations (e.g., in front of, inside, hole in, etc.). A system for this type of line drawing analysis has been proposed by one of us [27], and work on this system is in progress. It is hoped that the development of such systems will facilitate our understanding of GPM's and their role in SCLMIO analysis.

References

1. Y. Shirai, Understanding Intensity Arrays, A. I. Memo 263, M.I.T., 1972.

2. T. D. Garvey and J. M. Tenenbaum, On the Automatic Generation of Programs for Locating Objects in Office Scenes, IJCPR 2, 162-168, 1974.

3. M. Minsky, A Framework for Representing Knowledge, A. I. Memo 306, M.I.T., 1974.

4. M. Minsky and S. Papert, Progress Report, A. I. Memo 252, M.I.T., 1972.

5. T. Winograd, UNDERSTANDING NATURAL LANGUAGE, New York: Academic Press, 1972.

6. C. Hewitt, P. Bishop, and R. Steiner, A Universal Modular ACTOR Formalism for Artificial Intelligence, IJCAI 3, 235-245, 1973.

7. Z. Pylyshyn, What the Mind's Eye Tells the Mind's Brain, PSYCHOLOGICAL BULLETIN 80, 1-24, 1973.

8. C. Hewitt, Description and Theoretical Analysis of PLANNER: A Language for Proving Theorems and Manipulating Models in a Robot, AI Memo 251, M.I.T., 1972.

9. D. McDermott and G. Sussman, The Conniver Reference Manual, A. I. Memo 259, M.I.T., 1972.

10. R. N. Haber, The Nature of the Effect of Set on Perception, PSYCHOLOGICAL REVIEW 73, 335-350, 1966.

11. T. Cornsweet, VISUAL PERCEPTION, New York: Academic Press, 1970.

12. F. Attneave, Some Informational Aspects of Visual Perception, PSYCHOLOGICAL REVIEW 61, 183-193, 1954.

13. H. Barlow, R. Narasimhan, and A. Rosenfeld, Visual Pattern Analysis in Machines and Animals, SCIENCE 177, 567-575, 1972.

14. B. Julesz, Visual Pattern Discrimination, IRET-IT, 8, 82-92, 1962.

15. J.L.Muerle and D. Allen, Experimental Evaluation of Techniques for Automatic Segmentation of Objects in a Complex Scene, PICTORIAL PATTERN RECOGNITION (ed. Cheng, Ledley, Pollock, Rosenfeld), Washington, D. C: Thompson Book Co., 3-14, 1969.

16. A. Rosenfeld and M. Thurston, Edge and Curve Detection for Visual Scene

Analysis, IEEET-C 25, 609-615, 1971.

17. K. Koffka, PRINCIPLES OF GESTALT PSYCHOLOGY, New York: Harcourt, Brace, and World Inc., 1935.

18. M. Wertheimer, Untersuchungen zur Lehre von der Gestalt, PSYCHOL. FORSCH. 4, 301-350, 1923.

19. J. Beck, Similarity Grouping and Peripheral Discriminability Under Uncertainty, AMERICAN JOURNAL OF PSYCHOLOGY 85, 1-19, 1972.

20. S. Zucker, On The Foundations of Texture: A Transformational Approach, Tech. Rept. 331, Computer Science Center, University of Maryland, 1974.

21. N. Weisstein and C. S. Harris, Visual Detection of Line Segments: An Object-Superiority Effect, SCIENCE 186, 752, 1974.

22. J. Hochberg, PERCEPTION, Englewood Cliffs: Prentice-Hall, 1964, Chapter 5.

23. J. S. Bruner, On Perceptual Readiness, PSYCHOLOGICAL REVIEW 64, 123-152, 1957.

24. U. Neisser, COGNITIVE PSYCHOLOGY, New York: Appleton-Century Croft, 1967.

25. J. Hochberg, In The Mind's Eye, in R. N. Haber (ed.) CONTEMPORARY THEORY AND RESEARCH IN VISUAL PERCEPTION, New York: Holt, Rinehart and Winston, 1968.

26. M. Kelly, Edge Detection by Computer Using Planning, MACHINE INTELLIGENCE VI, Edinburgh University Press, 397-409, 1971.

27. L. S. Davis, Line Drawing Analysis: A Proposal, Tech. Rept. 334, Computer Science Center, University of Maryland, Sept. 1974.