

The Use of Geons for Generic 3-D Object Recognition

Sven J. Dickinson (Chair)

Dept. of Computer Science
University of Toronto
6 King's College Rd.
Toronto, Ontario
Canada M5S 1A4

Robert Bergevin

Dept. of Electrical
Engineering
Laval University
Sainte-Foy (Quebec)
Canada G1K 7P4

Irving Biederman

Dept. of Psychology
University of Southern
California
Los Angeles, CA
USA 90089-2520

Jan-Olof Eklundh

Computational Vision and
Active Perception Laboratory
Dept. of Numerical Analysis
and Computing Science
Royal Institute of Technology
S-100 44 Stockholm, Sweden

Roger Munck-Fairwood

Dept. of Electrical and
Electrical Engineering
University of Surrey
Guildford, Surrey
UK GU2 5XH

Alex Pentland

Vision and Modeling Group
Media Laboratory
Massachusetts Institute
of Technology
Cambridge, MA
USA 02139

Abstract

Biederman's introduction of geons to the vision community [BIE85] has spawned considerable interest in building geon-based vision systems. However, numerous issues must be addressed before such systems can make a practical contribution to machine vision. This paper collects together a number of distinguished researchers, each of whom has worked with geon-based recognition, in an effort to identify and discuss these issues.

1 Introduction (Chair)

In 1985, Biederman introduced to the computer vision community a theory of human visual object recognition called Recognition by Components (RBC) [BIE85]. The main tenets of RBC, supported by psychophysical experiments [BIE91; BIE93], are that: (i) the basic-level representation recovered from an image during *primal access* is metrically invariant, i.e., invariant to scale, translation, or orientation in depth, and (ii) the basic-level representation is composed of *parts* and their interrelations. In addressing these two principles, RBC also suggests that from nonaccidental contour relations in the image, a set of contrastive dichotomous (e.g., straight vs. curved axis) and trichotomous (e.g., constant vs. tapering vs. expanding/contracting cross-sectional sweep) 3-D volumetric primitive properties can be determined. The Cartesian product of the values of these properties gives rise to a set of volumetric primitives, or parts, which Biederman called *geons*. The notion of quickly recovering a set of object-centered, qualitative parts (geons) from an

image, both without a priori knowledge of the scene contents and without expending any effort on depth recovery, quickly attracted the attention of the object recognition community. Since then, a number of researchers from around the world have engaged in building computational models based on Biederman's RBC theory.

Beginning in 1988, Bergevin and Levine [BER88; BER89a] introduced their PARVO system, representing the first significant effort to build a vision system based on geons. From manually segmented line drawings (and from range data in 1991 [BER91]), they were able to recover and recognize a variety of multi-part objects. Then in 1989, Dickinson, Pentland, and Rosenfeld [DIC89] introduced a hybrid object representation integrating object-centered (volumes) and viewer-centered (aspects) models, with the object-centered models composed of parts closely resembling geons. This was to become the backbone of the OPTICA system which recognizes 3-D objects from real images [DIC92a; DIC92b].

A long-time advocate of recognition-by-parts, Pentland was the first to introduce parametric part representations to computer vision, using deformable implicit functions such as superquadrics [PEN86]. Similar to geons but with continuous rather than qualitative properties, this type of representation has been successfully used for recognition in segmented data [PEN91], and in modeling complex natural objects [SCL91].

In 1991, Munck-Fairwood investigated a logic programming approach to the recovery of geons from a manually segmented line drawing [MUN91a]. Recently, he has integrated causal probabilistic networks into his approach, assigning confidence measures to Biederman's dichotomous and trichotomous properties

[MUN91b]. Diverging from the analysis of labeled contours, Raja and Jain explored geon recovery from range data [RAJ92]. Their technique, which recovers geons from superquadrics fitted to range data, has been successfully tested on range images of real geon-based objects with straight axes.

Eklundh and his colleagues are building an active vision system whose goal is to recognize geon-based objects from images acquired by their stereo head [EKL90]. In recent work, Eklundh and Olofsson were the first to study the problem of geon-based object maintenance over time [EKL92], a critical problem facing the design and construction of geon-based active vision systems.

During these efforts by the vision community to build geon-based recognition systems, Biederman was exploring his own computational models for geon recovery. In 1988, Hummel, Biederman, Gerhardstein, and Hilton began exploring connectionist models for geon recovery, while in recent work [HUM92], Hummel and Biederman report a neural net model for shape recognition, using dynamic binding, that can recognize simple geon-based objects from line drawings.

Despite the growing interest in geons in the vision community, the above systems are still being applied to very constrained domains in which segmentation is often performed manually, lighting is tightly controlled, or the objects themselves are unrealistically simple. Nevertheless, the motivation for building generic recognition systems is strong. If a robotic vision system is to move about its environment, quickly recognizing new object exemplars from a large database of generic, articulated object models, as well as learning new object descriptions, it must have the capability to recover and match image representations which capture the coarse shape of an object in terms of its parts. Geons have emerged as the most popular shape representation for generic object recognition. However, before successful geon-based recognition systems can be realized, a number of important issues must be addressed, including:

- Are geons an appropriate modeling vocabulary for 3-D generic object recognition systems? If so, for what classes of objects are they best suited? If not, can the representation be altered or enhanced to make them more usable?
- What are the major obstacles in recovering geons from real imagery? What image features should be extracted in order to recover geons?
- What kinds of architectures and control strategies are appropriate for geon-based recognition?
- Finally, with the growing interest in active vision, what issues face the design and construction of geon-based active recognition systems? And what role can geon-based recognition play in a larger context where object manipulation may be required?

In the following sections, six researchers, representing broad experience in the design and construction of geon-based object recognition systems, address the above issues.

2 Testing the RBC Theory: An Approach Respecting the General Principles and the Computational Hypotheses

Robert Bergevin
Laval University

We see the original RBC proposal by Biederman as made up of two loosely connected sets of predictions: (i) the general principles, and (ii) the computational hypotheses. The general principles state that:

1. A line drawing which represents the discontinuities in an image is an efficient description of the image and is sufficient for the primal access task.
2. Objects are better represented and analyzed by decomposing them into their natural components (parts).
3. A qualitative description of the components is necessary and sufficient to permit fast access to a large database of object models.
4. Non-accidental instances of viewpoint-invariant features in the 2D line drawing are sufficient to permit fast access to the qualitative model of a 3D object.
5. Primal access for visual object recognition is obtained by matching a description of the spatial structure of components making up an object to an indexed database of models in a similar representation.

On the other hand, the computational hypotheses state that:

1. Five specific classes of 2D line groupings are sufficient to access the parts representation.
2. Segmentation should happen at concavities in the outline of an object.
3. The geons, a subset of the generalized cylinders, form an efficient qualitative shape representation for the parts which is suitable for the primal access task.
4. The symbolic descriptions for objects and models should include geon labels, aspect ratios, and relative sizes of parts.

While the psychological results directly support the general principles, the computational hypotheses are more like extrapolations or predictions regarding an actual implementation. This was meant to fill the gap between the levels of study in the psychological and computational perception fields (the "what" and the "how" of the problem, respectively).

The main assumption of RBC is the suitability of single view line drawings for primal access. This, together with the small set of volumetric primitives making up the geons, represent, in our opinion, the atomic kernel of the theory without which the goal of rapidly recognizing arbitrary unexpected objects cannot be attained. Of course, it could make sense to diverge from these principles and, say, analyze multiple 2D views of an object, recover the object from range data, reconstruct quantitative surfaces, volumes, or exact models, etc. But the

result would neither represent primal access nor even generic recognition, which is simply primal access followed by sophisticated high-level knowledge-based reasoning.

The difficulty to extract a very good line drawing in an uncontrolled imaging situation is well acknowledged in the computer vision field. Two approaches have been taken around this problem. Most groups assumed the (usually ideal) line drawing as given. Others decided to extract line drawings from real data (usually from range images) but had to restrict themselves to ideally shaped custom made objects. From the point of view of low level vision processes, the real question to be answered is how one goes from imperfect data from a non-ideally shaped object to the much simplified and compressed description in terms of an ideal line drawing. No one has yet a satisfactory answer to this question. Even worse, no one knows what would be such an ideal line drawing suitable for primal access and eventually generic recognition.

In our work with the PARVO system, two main goals were pursued [BER89a; BER92c]. First, we aimed to determine what are the underlying assumptions about the geometric properties of objects and their line drawings that make them acceptable to the kind of processing proposed in RBC. That is, what objects may be recognized with such an approach, what is their complexity, etc. Second, given such acceptable line drawings (ideal or noisy), we wanted to develop methods to actually perform the feature extraction, part segmentation, geon labeling, object description, and model matching to show that complex object recognition is actually possible in realistic conditions. Our main conclusions are that:

1. Primal access geon-based recognition is possible for a large number of man-made objects.
2. Recovery of geon-based descriptions from natural objects is likely to be very difficult.
3. To get acceptable line drawings will also be a quite difficult task especially if some of the object parts and connections do not respect the implicit assumptions in space or in the projected view.
4. Explicit part segmentation is needed and may still be possible even if some elements of the object or line drawing do not respect the implicit assumptions.
5. Face extraction and labeling is necessary for proper geon labeling.
6. Errors arising from blind labeling of the parts (without explicit verification of the hypotheses) is likely to destroy the indexing power of the geons.
7. Explicit verification is also necessary to make the recognition system robust against noisy (missing or spurious) data.
8. Parallel architectures and algorithms are necessary to cope with the combinatorial complexity arising from manipulating sets of features.

3 Recognition-by-Geons: Current Progress and Current Challenges

Irving Biederman

Peter C. Gerhardstein

Eric E. Cooper

Jozsef Fiser

University of Southern California

John E. Hummel

University of California, Los Angeles (UCLA)

In a fraction of a second humans are able to comprehend novel images of objects and scenes. Recognition-by-Components ([BIE87]) offered an account of this phenomenon characterized by four general assumptions: a) Objects are represented as an arrangement of simple convex or singly concave parts (*geons*), b) The geons can be distinguished by binary contrasts (differences) in viewpoint invariant properties (VIPs), such as straight vs. curved, rather than metric properties such as aspect ratio, c) The relations among the geons are explicit, such as PERPENDICULAR-TO or TOP-OF, as part of a structural description, rather than implicit in a coordinate space, and d) A relatively small number of geons is sufficient.

Results from priming experiments [BIE91; BIE93], in which the speed and accuracy of perceiving a briefly presented picture is facilitated by a prior presentation of a related picture, provide support for a simple, parts-based representation that is intermediate between a) local features such as filter outputs or image edges, and b) a global model (shape or concept) of the object. Moreover, variations in VIPs, such as changing a geon class from a cylinder to a brick, have a more significant effect on an object's real-time classification than variations in the geon's aspect ratio. A recent study of a successful face recognition system, which assumes a coordinate space and no intermediate primitives, showed fairly good overall accuracy in the recognition of images but its performance was strikingly different from that shown by human observers.

A neural net implementation of RBC ([HUM92]) takes a line drawing of an object as input and recruits a unit that is invariant over translation, size, and rotation-in-depth. The unit represents a structural description (geons + relations) of the object. The model's capacity for structural description derives from its original solution to the binding problem of neural networks. The input layer consists of a lattice of orientation-tuned cells. Fast Enabling Links (FELs) between cells that are collinear, coterminating, or closely parallel cause those cells that are currently activated by one geon to fire in synchrony but out of phase with the cells activated by another geon. The synchronous activity in a single time slice is used to bind independent units representing an object's geons and its attributes (e.g., vertical) and relations to other geons (e.g., above). For example, given an image of a vertical cylinder above a horizontal brick, the units for cylinder, vertical, and above will be induced to fire in synchrony but out of phase with the units for brick, below, and horizontal. Limitations in the tempo-

ral bandwidth for establishing phase differences results in accidental synchrony causing a large number of elements to be perceived as texture unless irrelevant activity is suppressed by attention.

Geon representations are not suitable for representing objects that do not readily decompose into simple parts. Differences among such objects, however, tend not to be incorporated as boundaries between basic level objects nor do they evidence the viewpoint invariance shown by objects that are readily described as an arrangement of geons [BIE93]. A possible benefit from employing a neural net model is that it can offer some capacity for achieving recognition (through partial activation) in the face of the ambiguity often inherent in images of such objects.

Challenges

Are there a small number of visual primitives? We are investigating this issue in the context of aspect graphs. Are there a limited number of part types that would be distinguishable from all other part types over large regions of the viewing sphere generating the aspect graph? For example, a brick can be identified as such from almost any viewpoint. If the areas of the viewing sphere corresponding to each interpretation are converted to probabilities, then geons have the property of low uncertainty in the Shannon sense. Is this a good criterion for identifying part types for human recognition? One attractive feature of this conceptualization is that one can define geons to be the set of volumes that have minimal viewpoint uncertainty. Low viewpoint uncertainty, moreover, may yield optimal conditions for a self-organizing network to develop hidden units that function as geon detectors.

Can a theory such as RBC be fast enough? Can the mechanisms of contour binding and geon activation be completed quickly enough to account for 100 msec recognition times?

Ignoring small variations. We can often determine, without a great cost in processing difficulty, volumetric structure despite the presence of small variations in decoration, flanges, etc. How?

4 Geons: A Viable Approach for Computer Recognition of Generic Objects?

Jan-Olof Eklundh

The Royal Institute of Technology

Visual object recognition is a wide and unwieldy subject on which computer vision has made but limited progress. One reason for this may be that problems range from recognizing known objects to general categorization, and that a common framework therefore may be elusive. Considering recognition of generic objects using volumetric primitives, like geons, can in this perspective form an attractive middle ground. Although it does not encompass all thinkable types of objects, since some natural objects like bushes may not even be characterized by their shapes, it does provide a rich enough world to be of interest. Moreover, it is also known that geon

representations allow indexing and hence rapid recognition. Finally, even though categorization problems in their most general form can hardly be addressed if only shape is used, recognition by shape primitives does apply to some of them. Notwithstanding important advances in the past few years, there are several critical issues to be addressed before we know how far the geon-based approach will take us. I will discuss some of these and how I think they can be treated at least in a general sense.

The first issue concerns how geons can be recovered from imagery. Most high level work has in this field, like in other recognition work, been based on perfect or near-perfect line-drawings. This is certainly not a realistic way to go. However, geons are qualitative in nature, and could be derived by some shape abstraction principle, which fits or associates primitive shapes to data. Several schemes in this vein have been suggested in the literature. However, a crucial requirement is that the typical geon properties should be made explicit, that is properties like being flat or curved, growing or tapering etc. This is not generally the case for approaches based on fitting methods, and may require non-linear techniques or cause uniqueness problems. Another important aspect is the treatment of structures at different scales. Generally speaking I don't think we know how to handle these problems, but I do believe that there exist promising methods.

The second issue is about the use of depth information. In his work on human recognition, Biederman is using line-drawings only. I think one must be careful in drawing conclusions about machine vision approaches from that. The display situations are such that the subject sees the object clearly separated from other objects in the fore- and background. Unless qualitative depth is provided either from accommodation cues or from binocular or other cues, a machine vision system using images with a large depth of field is faced with difficulties. Truly, Lowe and others have shown that recognition can be done in static monocular images. However, I think that then the possibilities of performing rather complex types of reasoning are needed, which goes against the general advantages of geon-based recognition. In principle, I feel that the problem should and can be posed in the presence of qualitative depth information, which can be used for segmentation and disambiguation. On the other hand I think that having dense depth information from stereo reconstructions or range sensing defines a completely different problem, and also a problem that is mainly of a technical character.

The third issue concerns whether the full geons need to be recovered or if observed subparts can have enough indexing power. Biederman has convincingly demonstrated that partial information is sufficient. However, this is again in a context where you don't see a large scene collapsed in depth. If those conditions are valid, partial information may not be constraining enough and can lead to combinatorial problems. On the other hand, if it seems as if there is some other process selecting a region of interest, or generally guiding the attention, then there are viable computational techniques which could be based on typical partial information that we can ob-

tain from real images.

Generally I think that geon-based recognition of generic objects can be a powerful approach, even though it has limitations. However, I also think that the technique must be applied in a context where the issues I've listed, and also other problems, can be addressed appropriately.

5 Geons and the 'What'-function

Roger Munck-Fairwood

Li Du

University of Surrey

Generic object recognition (GOR) may be regarded as the attempt to fulfill the 'what'-function of machine vision. We have identified a fundamental question which must be answered if GOR (as it is currently understood) is to be posed as a scientific pursuit. The question is: How can the 'what'-function be defined for both human and machine vision?

We further propose a system architecture which addresses the first three of the issues listed in the introduction. We also discuss the fourth issue.

5.1 'What'-function and the Purpose Concept

In our definition, the 'what' function puts a 3D entity into a predefined bin of classification on the basis of the recovered geometrical properties. Let us assume that W is the vocabulary for describing the geometry of 3D entities (as deliverable by a collection of lower level vision processes), C is the set of predefined classification bins and M is the classifications process. A single task of recognition can then be formally expressed as $c = M(w)$ where $c \in C$ and $w \in W$.

In the light of this formalism, the current talk about generic object recognition systems which should work 'reliably' implies the assumption that there exists a universal (all-purpose) C . Clearly, this is a false assumption, because even given a perfect w (geometrical description of an object), the way the bins are designed strongly depends on either engineering criteria or cultural customs. Regarding W , one can always resort to object geometry, in mathematical terms, as an ultimate unambiguous object description (but *not* classification), although the lower level processes might not yet be able to provide this from an image. Whereas for C , we do not even know how many overlapping C 's exist in our culture.

Another apparent flaw is that in the absence of a constraint on C , the indexing power of a GOR system can be grossly deceptive. Given a W (e.g. geon clusters), it is a trivial task to imagine an M which results in a C with hundreds of thousands of bins but many or most of these may be practically meaningless.

In conclusion, this formalism pinpoints two fundamental aspects of GOR: (i) The bins of C which must satisfy a purpose or a set of purposes and (ii) an architecture of geometry recovery processes which delivers the best collection of W . The benefit of this formalism is that the generality (or 'generic-ness') of any 'generic object

recognition system' can be *measured* – not by the number of bins in a C – but by the richness of the set of purposes P which its C satisfies and the versatility of its W which can be delivered by the recovery sub-system. So, we no longer regard GOR as universal but bounded by purposes on the top end and richness of geometric description on the lower end.

5.2 A Framework for Recovery Sub-system

We have outlined [MUN93] a recovery sub-system design to provide a versatile repertoire of image labeling tools as a *servant* to an object classification system. The basic principles behind the system design include the following. (a) It is a servant to a classification sub-system having a high degree of transparency in the sense of making process parameters (scale, thresholds, etc.) explicit. (b) The architecture has the potential to incorporate any existing approaches of 3D geometry recovery. (c) The image labeling vocabulary is divided both into levels of abstraction and levels of cost. (d) Reasoning occurs in three directions: bottom-up (b-u), top-down (t-d) and horizontally. When b-u and t-d are in conflict, either b-u can dominate (corresponding to visual illusion in human vision) or t-d can dominate (corresponding to subjective contours).

We illustrate the system using 2D edge-based features, but surface features can also be incorporated as an alternative channel to the recovery of primitives (geons being the best so far).

5.3 Object Classification Sub-system

This sub-system has two-way communication with the W -sub-system. It not only takes b-u input from the latter but gives cognitive (predictive) information to it when possible. Its basic function is to take the recovered geometry at the geon-cluster level of the W -system and put each cluster into the appropriate bin of C .

In a given engineering task, the human operator may define the bins of C , thus simplifying the task of this sub-system. But in general, the sub-system should automatically create the bins of C according to a given purpose or set of purposes P .

5.4 About Active Vision

Our definition of GOR shares the 'purpose' concept with that used in the study of active vision, but from an entirely different point of view. Currently in active vision research, the emphasis is shifted towards sensor-motor interaction and the 'purpose' concept is used to circumvent the need for the 'what'-function (e.g. by gazing on a bright spot or a color pattern). In contrast, we employ the 'purpose' concept to directly address the issue of object recognition by linking P with C . We consider object recognition as a key component for a general active vision system, where often the actions can only be taken when 'what' is answered.

6 Parts Representations

Alex Pentland

Massachusetts Institute of Technology

The representation of objects by their parts has a long tradition in computer-aided design, simulation, and cognitive psychology. Indeed, in these areas it is the dominant strategy for representing complex 3-D objects. It is absolutely clear, therefore, that part representations are excellent for many computational and cognitive tasks. What is not so clear is how they might be useful in computer vision.

The first parts representation was suggested by Binford 1971; this is the idea of generalized cylinders (G.C.). Because of the relatively unconstrained nature of this representation, the recovery process seems to require elaborate line grouping and reasoning. Consequently there are relatively few reports of recovering such descriptions from real imagery. Moreover, because such descriptions are often not unique, it is unclear how they aid in object recognition.

About 1985, these problems with G.C. representations began to be taken seriously, and two different proposals arose to deal with them. One variation on the G.C. idea is due to Biederman [BIE85], who suggested using the Cartesian product of qualitative properties such as tapering, cross-section, etc., in order to create a qualitative taxonomy of G.C. One advantage of this type of representation, called *geons*, is that the properties can be chosen to be ones that are more easily recovered from imagery. Another is that it provides a way to define qualitative shape classes, an important problem in general-purpose vision.

To date, however, only Dickinson, Pentland, and Rosenfeld [DIC92b] have successfully used this approach to recognize objects in real imagery, although Bergevin and Levine [BER89a], Hummel and Biederman [HUM92], and Munck-Fairwood [MUN91b] have reported good success in interpreting line drawings. The principal problems with this approach seem to be the difficulty in extracting sufficiently good line drawings, and the idealized nature of the *geon* representation.

A different variation of the G.C. idea was suggested by Pentland [PEN86], who proposed a parametric version of G.C. based on deformable implicit functions. The original version of this proposal used superquadric functions with multiplicative deformations; later versions used spheres with additive ("Modal") deformations. Use of a parameterized implicit function (deformed superquadrics or spheres) converts the problem of recovering a description into a relatively simple numerical optimization. Moreover, if the parameterization is orthogonal then the description is unique, making the recognition problem much easier.

In recent years, several authors have reported success at recovering this type of description and using it for recognition, e.g., Pentland and Sclaroff [PEN91] were able to recognize people's head shape with 96% accuracy. The major limitations of this approach seem to be the requirement for pre-existing part segmentation, and an initial estimate of orientation.

It is interesting to compare the *geon* and parametric approaches. Both describe a similar range of shapes; perhaps the major difference is that one is qualitative while the other is quantitative. This similarity has made

it possible to generate *geon* statistics using parametric models [DIC92a], and allowed Raja and Jain [RAJ92] to use superquadrics to recover *geons*.

It seems to me that the qualitative *geon* representation is best suited to understanding intersections and groupings of parts, while the parametric representation is best for classifying individual part shapes. I speculate that the most successful systems will employ first a qualitative pruning/grouping stage to perform segmentation, and then a parametric fitting stage to extract detailed part shape. This approach was recently successfully employed in [DIC92c].

A final question is representational adequacy. It seems clear that to model natural objects the G.C. representation must be augmented by a displacement map as in [SCL91]. Moreover, the G.C. representations currently used in vision are much simpler than those used in CAD. Consequently, I predict that the more expressive G.C. representations [PEN91][SCL91], will become increasingly popular.

7 The Obstacles Facing Practical Geon-Based Recognition

Sven Dickinson

University of Toronto

I see three major obstacles facing the success of *geon*-based object recognition systems: (i) the recovery of *geons* from real imagery; (ii) the difficulty in *explicitly* modeling real objects using *geons*; and (iii) the lack of representational power, e.g., depth and fine shape, provided by *geons* for the task of interacting with the world.

7.1 Recovery

In our work on *geon*-based recognition, we have attempted to address the issues in building *geon*-based systems that work on real imagery [DIC92a; DIC92b]. We chose an approach to volumetric primitive recovery based on part-based aspect matching which, in fact, can be applied not only to *geons*, but to other classes of volumetric primitives as well. From our analysis, we discovered that image regions, not lines or line groups, were the most appropriate features to extract from an image. Image regions not only provide a natural grouping of image contours, i.e., the bounding contours of a region, but region grouping provides a powerful inferring mechanism for volume recovery. Our *aspect hierarchy*, a multi-level, viewer-centered representation of a set of volumetric parts, allows us to recover *geons* in the presence of noise, occlusion, and low probability views.

Despite the utility of regions in recovering *geons*, we are still faced with the problem of quickly extracting regions from an image. In addition, how do we deal with the inevitable under- and over-segmentation of image regions? And finally, at what stage in the recovery of *geons*, do we attempt to rectify such segmentation problems: immediately following segmentation? at the *geon* verification stage? or, at the object verification stage? In working with real images, it becomes very clear that the segmentation cannot be decoupled from the recogni-

tion. I believe that a dynamic feedback scheme between the various stages in the recognition process must be explored before a geon-based object recognition system can be successfully applied to real images. We are currently exploring feedback schemes applied to segmentation and recognition.

7.2 Modeling

One of the most challenging tasks in building a geon-based recognition system was finding geon-based objects to test it on! In fact, we ended up constructing our own geons in order to build our own objects. That is not to say that there are not geon-based objects in the world. However, what we found was that although many of the real objects we did look at could be coarsely described in terms of geons, there was a plethora of detail in the objects that did not contribute to their coarse shape. Even ignoring surface markings and texture, there would be a lip on a cup, a small ridge around the edge of a table, or other fine object structure whose projected contours in the image were quite salient. Although Biederman suggested the use of nonaccidental relations between contours in recovering geons, neither he nor the line-drawing based geon-recovery systems which he inspired addressed the issue of structural scale. I believe that geons are an appropriate indexing primitive for recognition. However, we must be able to recover abstract geon descriptions of objects with structural detail.

7.3 Representational Power

For a vision system to interact with the world requires that it not only identify or recognize the objects that it wishes to interact with, but that is sufficiently localize (in depth) and characterize (in dimension) the objects in order to manipulate them. Although geon-based object recognition techniques may prove suitable for the identification task, they fall short in localizing and characterizing the object. Although geons were not intended for this task, we are studying ways in which geon-based recognition can be used within a larger framework.

At the University of Toronto, we are building a vision-based mobile robot whose goal is to allow handicapped children to interact with their environment through the use of a robotic manipulator. The child interacts with the system by instructing it to find and manipulate toys on a table; hence, we need both recognition and localization of objects. In the spirit of decoupling the "what" and "where" tasks, which are typically tightly coupled in verification-based object recognition approaches, we can maintain the speed and efficiency of a qualitative front-end recognition module, and invoke a localization module only when necessary. We have built such a system, using qualitative shape to constrain the fitting of physics-based deformable models to the image [DIC92c].

References

- [BER88] R. Bergevin and M. D. Levine, "Recognition of 3-D Objects in 2-D Line Drawings: An Approach Based on Geons", Technical Report TR-CIM-88, Nov. 1988.
- [BER89a] R. Bergevin and M. D. Levine, "Generic Object Recognition: Building Coarse 3D Descriptions from Line Drawings", IEEE Computer Society Workshop on Interpretation of 3D Scenes, November 27-29, Austin, Texas, 1989.
- [BER91] M. D. Levine, R. Bergevin, and Q. L. Nguyen, "Shape Description Using Geons as 3D Primitives," International Forum on Visual Form, May 27-30, Capri, Italy, 1991.
- [BER92c] R. Bergevin and M. D. Levine, "Generic Object Recognition: Building and Matching Coarse Descriptions from Line Drawings," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 1, January 1993, pp 19-36.
- [BIE85] Biederman, I., "Human Image Understanding: Recent Research and a Theory", *Computer Vision, Graphics, and Image Processing*, Vol. 32, 1985, pp 29-73.
- [BIE87] Biederman, I., "Recognition-by-Components: A Theory of Human Image Understanding", *Psychological Review*, 94, pp 115-147.
- [BIE91] Biederman, I. and E. E. Cooper, "Priming contour-deleted images: Evidence for intermediate representations in visual object recognition", *Cognitive Psychology*, 23, pp 393-419.
- [BIE93] Biederman, I., "Recognizing Depth-Rotated Objects: Evidence for 3D Viewpoint Invariance", *Journal of Experimental Psychology: Human Perception and Performance*, in press.
- [DIC89] Dickinson, S. J., A. P. Pentland, and A. Rosenfeld, "A Representation for Qualitative 3-D Object Recognition Integrating Object-Centered and Viewer-Centered Models", Technical Report CAR-TR-453, Center for Automation Research, University of Maryland, June, 1989. (also appears with the same title in: K.N. Leibovic (ed.), *Vision: A Convergence of Disciplines*, Springer Verlag, NY, 1990, pp 398-421.)
- [DIC92a] Dickinson, S. J., A. P. Pentland, and A. Rosenfeld, "3-D Shape Recovery using Distributed Aspect Matching", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, special issue on Interpretation of 3-D Scenes, Vol. 14, No. 2, February 1992.
- [DIC92b] Dickinson, S. J., A. P. Pentland, and A. Rosenfeld, "From Volumes to Views: An Approach to 3-D Object Recognition" *Computer Vision, Graphics, and Image Processing: Image Understanding*, special issue on CAD-Based vision, Vol. 55, No. 2, March 1992.
- [DIC92c] Dickinson, S. J. and D. Metaxas, "Using Qualitative Shape to Constrain Deformable Model Fitting", Sensor Fusion V, SPIE OE/Technology '92, Boston, MA, November, 1992.
- [EKL90] Eklundh, J., M. Li, and G. Olofsson, "Representations for 3-D Geometric Modeling", ESPRIT BRA-3038, Vision as Process, Internal Report IR.C.1.1., May, 1990.
- [EKL92] Eklundh, J. and G. Olofsson, "Geon-Based Recognition in an Active Vision System", in: ESPRIT-BRA 3038, Vision as Process, Springer-Verlag ESPRIT Series, (to appear).
- [HUM92] Hummel, J. E. and I. Biederman, "Dynamic Binding in a Neural Net Model for Shape Recognition", *Psychological Review*, 99, 1992, pp 480-517.
- [MUN91a] Fairwood, R. C. (now R. C. Munck-Fairwood), "Recognition of Generic Components Using Logic-Program Relations of Image Contours", *Image & Vision Computing*, Vol 9 (2), pp 113-122, April 1991.
- [MUN91b] Fairwood, R. C. (now R. C. Munck-Fairwood) and G. Barreau, "A Belief Network for the Recognition of 3-D Geometric Primitives", 6th International Conf. on Image Analysis and Processing, Como, Italy, September 1991.
- [MUN93] Du L., and Munck-Fairwood, R. C., "A Formal Definition and Framework for Generic Object Recognition", To appear in the Proceedings of the 8th Scandinavia Conference on Image Analysis, Tromso, Norway, May, 1993.
- [PEN86] Pentland, A. P., "Perceptual Organization and the Representation of Natural Form", *Artificial Intelligence*, Vol. 28, 1986, pp 293-331.
- [PEN91] Pentland, A. and Sclaroff, S., "Closed-Form Solutions For Physically-Based Modeling and Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, No. 7, pp 715-730.
- [RAJ92] Raja, N. S. and A. K. Jain, "Recognizing Geons from Superquadrics Fitted to Range Data", *Image and Vision Computing*, Vol. 12, No. 3, April 1992, pp 179-189.
- [SCL91] Sclaroff, S., and Pentland, A., "Generalized Implicit Functions for Computer Graphics", *ACM Computer Graphics*, Vol. 25, No. 2, pp 247-250.