Object Recognition in the Geometric Era: a Retrospective

Joseph L. Mundy

Division of Engineering,
Brown University
Providence, Rhode Island
mundy@lems.brown.edu

Abstract. Recent advances in object recognition have emphasized the integration of intensity-derived features such as affine patches with associated geometric constraints leading to impressive performance in complex scenes. Over the four previous decades, the central paradigm of recognition was based on formal geometric object descriptions with a focus on the properties of such descriptions under perspective image formation. This paper will review the key advances of the geometric era and investigate the underlying causes of the movement away from formal geometry and prior models towards the use of statistical learning methods based on appearance features.

1 Introduction

Object recognition by computer has been an active area of research for nearly five decades. For much of that time, the approach has been dominated by the discovery of analytic representations (models) of objects that can be used to predict the appearance of an object under any viewpoint and under any conditions of illumination and partial occlusion. The expectation is that ultimately a representation will be discovered that can model the appearance of broad object categories and in accordance with the human conceptual framework so that the computer can "tell" what it is seeing.

Advantages of geometric description From the earliest attempts at recognition, geometric representations have dominated the development of the theory and resulting algorithms and systems. There are a number of reasons why geometry has played such a central role.

- Invariance to viewpoint - Geometric object descriptions allow the projected shape of an object to be accurately predicted under perspective projection.
- Invariance to illumination - recognizing geometric descriptions from images can be achieved using edge detection and geometric boundary segmentation. Such descriptions are reasonably invariant to illumination variations.

- Well developed theory - geometry has been under active investigation by mathematicians for thousands of years. The geometric framework has achieved a high degree of maturity and effective algorithms exist for analyzing and manipulating geometric structures.
- Man-made objects - a large fraction of manufactured objects are designed using computer-aided design (CAD) models and therefore are naturally described by primitive geometric elements, such as planes and spheres. More complex shapes are also represented with simple geometric descriptions, such as a triangular mesh or polynomial patches.

There are, of course, deficiencies of the geometric approach to recognition, but the discussion of such limitations will be postponed until after a review of the broad sweep of geometric recognition research over the last four decades.

2 The beginning

In the 1950s and early 1960s ideas from signal processing and detection theory, such as autocorrelation and template matching, were exploited to form the first object recognition systems. Much of the research focus was on 2-d pattern classification applications such as character recognition, fingerprint analysis and microscopic cell classification. These early decades were dominated by methods of statistical pattern recognition and perception classifiers based on parametric learning. Even so, the features used in these classification schemes were often derived from geometric descriptions. For example, an early approach [34] (1962) to the definition of features for character recognition was based on geometric invariance using moments. Geometric invariance will re-appear as a major research thrust in the early 1990s, three decades later. This example illustrates that recognition ideas are continually re-visited as computational power and feature segmentation methods advance.

2.1 The blocks world

The dependence on statistics and signal methods rapidly gave way to the theme of *artificial intelligence*, coined by Marvin Minsky and John McCarthy around 1956. The new approach focussed on establishing a theoretical framework for cognitive tasks, such as vision, where computers could carry out the necessary reasoning using formal logic and other mathematical tools. The plan was to start with a simplification of the world so that the mathematical models can apply rigorously and to solve the resulting recognition problem completely before proceeding to more difficult situations.

For the computer vision problem, this simplification is called *the blocks world* where objects are restricted to polyhedral shapes on a uniform background. Polyhedra have simple and easily represented geometry and the projection of polyhedra into images under perspective can be straightforwardly modeled with a projective transformation. Under this projection, lines in 3-d map to lines in 2-d

and polyhedral faces project to polygons. The goal is to be able to recognize general polyhedral shapes in an arbitrary spatial arrangement including significant occlusion of one object by itself or others.

The blocks world framework dominated the vision research agenda for over a decade before it was abandoned to tackle more realistic scenes. It is not that all the problems of recognizing polyhedral objects and structures made up of polyhedra were definitively and completely solved. Instead it became clear that too many assumptions were being made in recognition strategies that could not be expected to hold in real world scenes. This tension between the desire for a sound theoretical basis for recognition and the ability to confront the complexities of recognizing complex objects such as trees and the human form, will re-immerge repeatedly during the geometric era.

2.2 Roberts and the blocks world

Perhaps the most complete and powerful recognition system of the blocks world was that of L. G. Roberts [64]. Roberts' recognition algorithm exhibited most of the steps that are still followed today, some four decades later. He carefully considered how polyhedra project into perspective images and established a generic library of polyhedral components that could be assembled into a composite structure. His philosophy towards recognition is defined by the quote, '... we shall assume that the objects seen could be constructed out of parts with which we are familiar. That is, either the whole object is a transformation (projection [1]) of a preconceived model, or else it can be broken into parts that are. ... The only requirement is that we have a complete description of the three-dimensional structure of each model.'

Roberts developed his own edge detector and line fitting algorithms along with feature grouping heuristics appropriate for polyhedral projections. The feature grouping formed hypotheses for 3-d polyhedral vertices and edges that were validated by solving for the associated projective camera model parameters. Interestingly, his linear resection algorithm is still used to initialize non-linear solvers in modern camera calibration methods. The result of these steps is shown in Figure 1 where the final extracted scene is displayed from a different viewpoint in order to demonstrate the accuracy and completeness of the recognition result.

The constraints of polyhedral scenes were exploited in many different ways including the powerful approach of constraint labeling initiated by Adolfo Guzmán [30] and fully exploited by David Waltz [81] and others [20, 35, 47]. In this work, the local constraints of the polyhedral vertices and edges can be propagated to neighboring vertices while ruling out multiple interpretations of the convexity and occluding state of projected boundaries. These ideas were later put on a fully algebraic basis by Kokichi Sugihara [76].

The culmination of the blocks world effort was the *MIT copy demo* [84]. The demo consisted of a robot observing a designed structure of polyhedral blocks

[1] Added for clarification within the quoted context

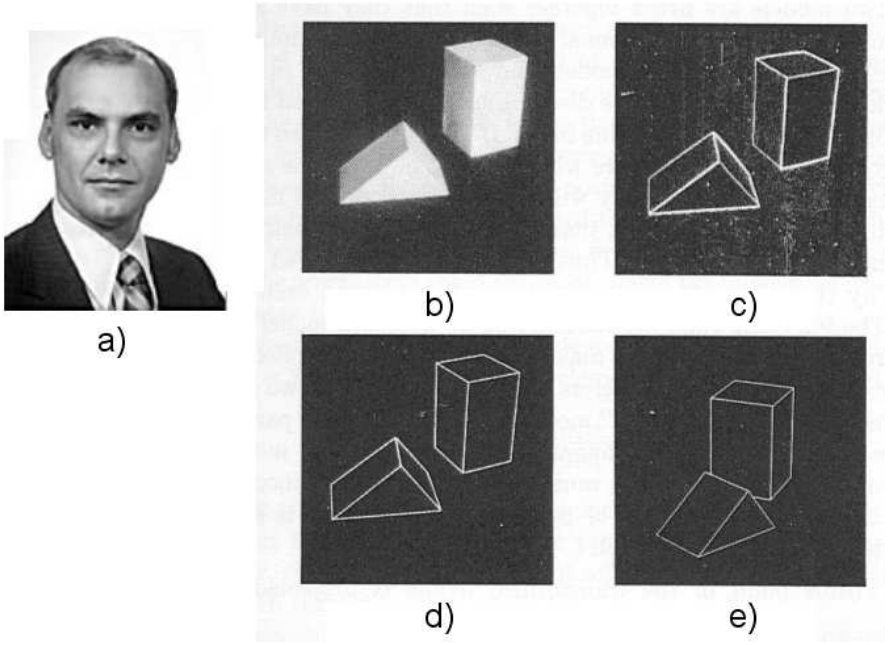


Fig. 1. A system for recognizing 3-d polyhedral scenes. a) L.G. Roberts. b)A blocks world scene. c)Detected edges using a 2x2 gradient operator. d) A 3-d polyhedral description of the scene, formed automatically from the single image. e) The 3-d scene displayed with a viewpoint different from the original image to demonstrate its accuracy and completeness. (b) - e) are taken from [64] with permission MIT Press.)

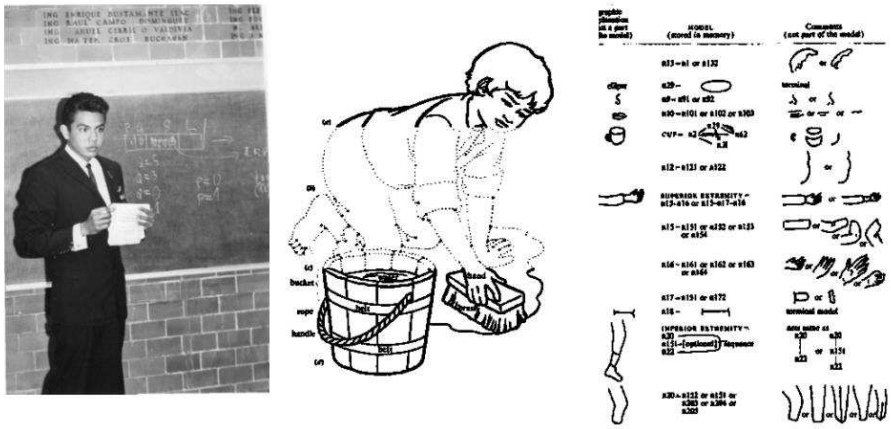
and then recreating a copy of the structure from a pile of unordered blocks. This task required recognition as well as an analysis of stability and hand-eye coordination. A similar achievement for a recognition system of the modern era does not come readily to mind.

What the blocks world didn't confront The blocks world avoided numerous difficulties such as:

- curved surfaces and boundaries;
- articulated and moving objects;
- occlusion by unknown shapes;
- complex background and 3-d texture such as foliage;
- specular or mutually illuminating surfaces;
- multiple light sources and remote shadowing;
- transparent or translucent surfaces.

The blocks world was extended in various ways to begin coping with these conditions. An early exploration of the issues that arise in the recognition of generic

curved objects was carried out by Guzmán [31]. His approach is illustrated in Figure 2. This work can be seen as an extension of the blocks world philoso-



a) b) c)

Fig. 2. A system for recognizing 2-d curved objects in line drawings. a) A. Guzmán in 1964. b) The feature analysis of a line drawing. c) A set of parts that can be used to describe generic curved objects. (b) and c) are taken from [31] with permission.)

phy. By restricting the problem to line drawings, many of the difficult scene rendering issues can be avoided and research can focus on what happens when curved surfaces intersect and occlude and where generic objects categories can exhibit a wide range of composite parts. For example, in Figure 2 c) there can be many types of pants legs, with and without creases and highly variable geometric relations between such parts.

In spite of this innovative use of parts and constraint relations to enable the recognition of objects in more real-world scenes, the restriction to ideal line drawings seemed too far away from the real vision problem to build to a major focus of the recognition community. Instead, a new geometric representation was discovered that offered a way to extend the blocks world to composite curved shapes in 3-d - the generalized cylinder.

3 Binford and the world of generalized cylinders

The next major advance in representations for recognition was the generalized cylinder (GC) originated by Thomas Binford [8]. The key insight is that many curved shapes can be expressed as a sweep of a variable cross section along a curved axis. Issues such as self-intersection and surface singularities do arise but

shapes like a coffee pot or cup are easily handled. An example of automatically extracting an object description using generalized cylinders is shown in Figure 3. This example was taken from the work of Gerald Agin [2], a Binford student at Stanford. Agin developed a structured light range camera and used generalized cylinders to model various curved shapes, such as dolls.

The recognition of simple curved 3-d objects, such as a hammer, based on the Agin range camera and generalized cylinder components was carried out at the same time by another Binford student, Ram Nevatia [56, 57]. Nevatia has maintained a long-term commitment to the generalized cylinder representation and has pursued recovery and recognition of GC objects from intensity images as a major research goal. An example of Nevatia's later work some two decades later on GC part decomposition for object recognition is shown in Figure 4 [85]. This result is quite an achievement given the relatively weak evidence for GC part boundaries and interfaces in the image.

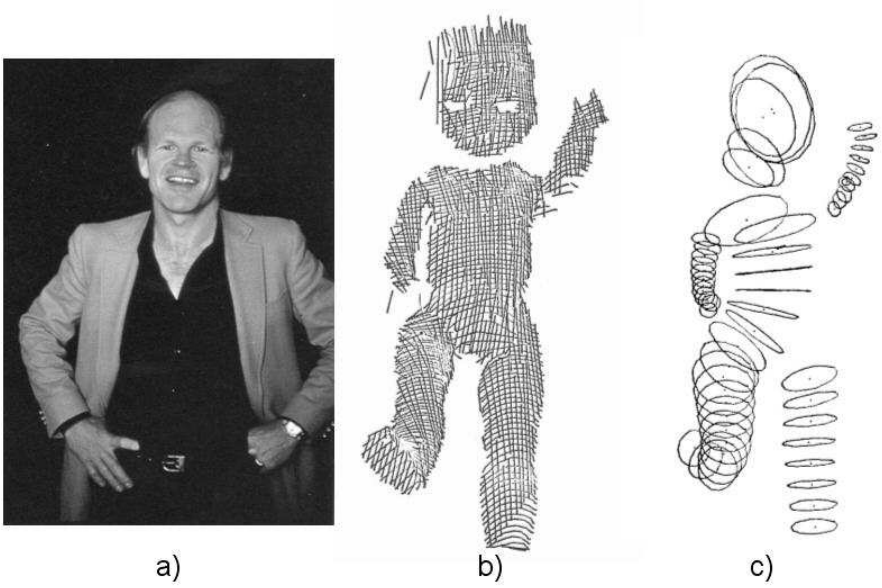


Fig. 3. The representation of objects by assemblies of generalized cylinders. a) Thomas Binford. b) A range image of a doll. c) The resulting set of generalized cylinders. (b) and c) are taken from Agin [1] with permission.)

3.1 ACRONYM

Another Binford student, Rodney Brooks, developed a recognition system based on symbolic geometric constraints on objects composed of GC parts [13]. The sys-

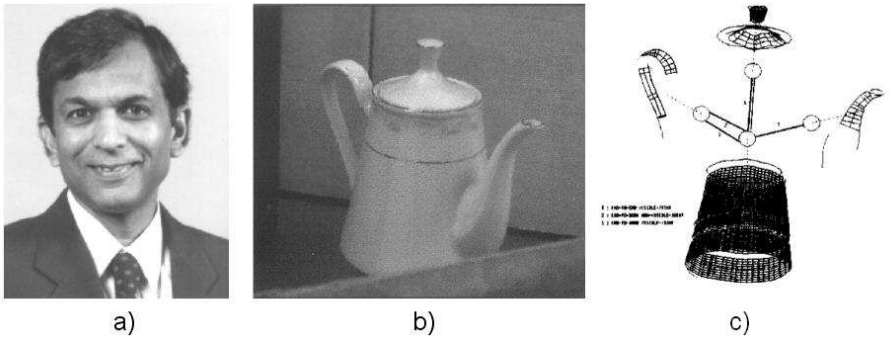


Fig. 4. Recognition by generalized cylinder parts. a) Ram Nevatia. b) An intensity image of a coffee pot. c) Automatically grouped and classified GC parts. (b) and c) are taken from [85] with permision.)

tem could essentially prove theorems concerning the existence of a parameterized GC configuration with associated tolerances. The system was called ACRONYM to avoid deriving a contrived name for the system, since ACRONYM is cleverly self-referential [2]. The Defense Advanced Projects Agency (DARPA) and the Cen-

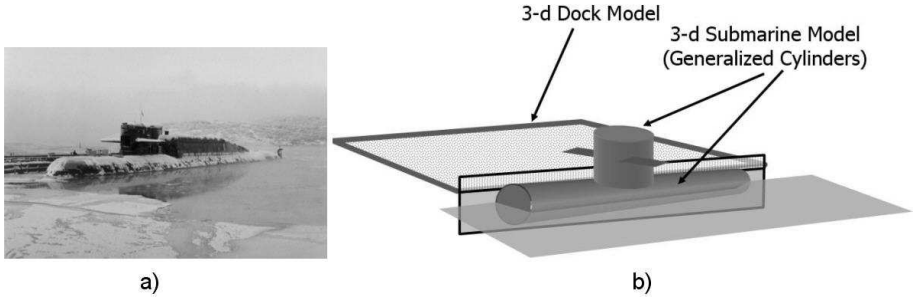


Fig. 5. The SCORPIUS project. a) A submarine at dock. b)An ACRONYM generalized cylinder model for the scene in a).

tral Intelligence Agency (CIA) established a classified project to use ACRONYM to recognize targets such as submarines as illustrated in Figure 5. The goal was to assist strategic intelligence analysts that monitor military installations using aerial photography. The project, called SCORPIUS, was designed to exploit var-

[2] Binford's next generation system was called SUCCESSOR [9], thus eliminating the need for any future acronyms.

ious parallel computing architectures developed by DARPA in conjunction with the Strategic Computing Program (1983-1993) [65]. Since the SCORPIUS program was classified, it is not clear how effectively the ACRONYM recognition system performed. The results must have been encouraging enough since a new project, called RADIUS, was launched in 1993 with similar application goals [25]. However, the emphasis of RADIUS was on change detection and automated 3-d modeling from imagery rather than recognition.

4 Aspects

The early period of object recognition research was based solidly on the premise that objects live in 3-d space and the 3-d structure can account for all the changes in appearance that arise from viewpoint changes. There was not much interest in explaining image intensity variations except for the early work by Horn [33]. The rationale was that objects can be recognized from their outlines and interior intensity discontinuity boundaries and that these features can be reliably recovered without requiring an in-depth understanding of reflectance and image intensity formation. This framework is known as object-centered representation.

An alternative representational scheme arose in the 1970s based on a network of the distinct 2-d views of an object, called an *aspect graph*. The pioneering work in this area was by Stephen Underwood and Clarence Coates [80], Jan Koenderink and Andrea Van Doorn [39] and Indranil Chakravarty [17]. A graphical representation of a set of 2-d views of a polyhedral shape is shown in Figure 6, as described in [80]. The idea of pre-compiling 2-d views into an efficient recognition plan was also developed by Chris Goad [27], who viewed recognition planning as a form of automatic computer programming. Repeated view calculations should be pre-compiled off-line to achieve high performance during recognition runtime processing. Later the computation of aspect graphs was extended to generalized cylinders by Jean Ponce and David Kriegman [41]. In general, the graph of related object views is called an *aspect* graph. The nodes of the graph represent object views that are adjacent to each other on the unit sphere of viewing directions but differ in some significant way. The most common view relationship in aspect graphs is based on the topological structure of the view, i.e., edges in the aspect graph arise from transitions in the graph structure relating vertices, edges and faces of the projected object.

The aspect graph representation gained a lot of momentum with resonance from the psycho-physics community where some researchers embraced the notion that human vision is view-based rather than object centered [77]. The hope was that visual aspects, compiled from 3-d models, or learned from example images could enable an efficient recognition strategy by guiding the search for image features. The family of deformable generalized cylinder parts called *geons* were introduced by Irving Biederman [7] who demonstrated that human object recognition can be characterized by the presence or absence of geons in the 3-d scene. Sven Dickinson, Sandy Pentland and Azriel Rosenfeld developed an aspect graph formulation of geon primitives for the recognition of 3-d objects [22].

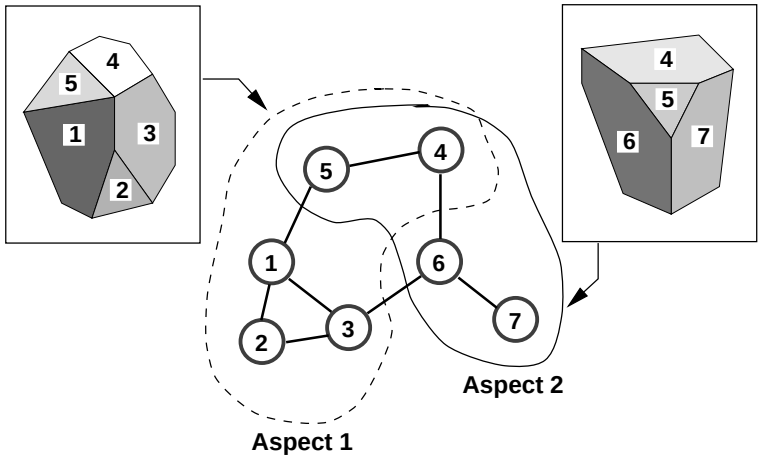


Fig. 6. Two views of a polyhedral solid. The adjacency of projected polygonal faces forms a graph. The view-based description is learned by associating new view structures with the existing graph. The figure is similar to one from [80].

The formal goal of precise computation of aspect graphs encountered some major difficulties in the 1990s. It was shown by Harry Plantinga and Charles Dyer [60] that under perspective viewing that the size of polyhedral aspect graphs can grow as rapidly as n^9. For curved surfaces, the complexity is dramatically greater. Sylvain Petitjean [59] found that the complexity of the aspect graph of algebraic surfaces is on the order of d^{18}, where d is the degree of the surface. This complexity arises since there are many small scale transitions that are topologically significant but may not be relevant for object recognition. Since the viewing distance is not known in advance, it is difficult to say what topological events are important and therefore the aspect graph enterprise becomes application specific.

The example of Figure 7 provides a clear illustration of this issue and was used in a debate heralding the end of substantial research on the formal aspect graph [23]. The dimples on the golf ball introduce intractable complexity to the graph representation but are not of individual significance in an effective description of the object class. More recently, Ben Kimia has formulated an aspect graph based on the geometric similarity of object views as measured by elastic deformation [21]. While this approach avoids the polynomial explosion of views based on topological details, the problem of scale still persists.

5 The era of pessimism

The early geometric period was founded on the notion that bottom-up boundary descriptions could be formed from single intensity views of an object. This

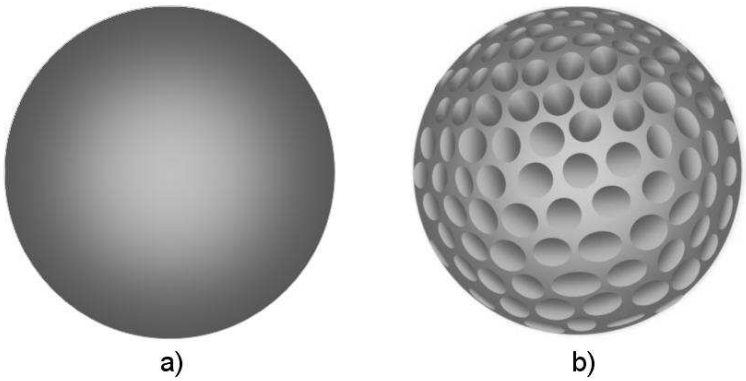


Fig. 7. The problem of scale for the aspect graph representation. a) A golf ball seen from a large viewing distance. b) The same ball from a close viewpoint. Each dimple generates a combinatorial explosion of occlusion events with respect to the other dimples.

process, later to be called *perceptual grouping* [48, 45, 69] presented some difficult problems such as:

- low contrast image intensity at boundaries;
- background clutter with high edge density;
- occlusion by objects with complex texture.

As an example of the first point, an image of a polyhedral edge will exhibit no intensity discontinuity at all if the illumination is directed along the direction of the mean surface normal of the intersecting planar faces (assuming Lambertian reflectance). This condition can be easily observed for polyhedral surfaces of modest complexity and thus reliable boundary detection cannot be practically achieved. The missing edges must be hypothesized based on reasoning about the object shape, which dictates that bottom-up grouping cannot be done in advance of considering a model hypothesis.

These difficulties generated a period of pessimism concerning the completeness and stability of bottom-up segmentation processes. Instead, a number of researchers implemented recognition systems based on fragmentary feature segmentations in terms of 2-d point and line or curve segments. The organization of these features is based on a specific individual object model rather than the generic descriptions that dominated the early period.

Some early examples of this approach can be seen in the 1970s [3] and [58]. A system for the recognition of 3-d parts with planar surfaces was developed by Walter Perkins at General Motors. The goal was the so-called "bin-picking" problem where the recognition process determined the pose (rotation and translation) of the object in a world coordinate frame so that the object could be

placed by a robot into a fixture for subsequent manufacturing operations. An example of part recognition is shown in Figure 8.

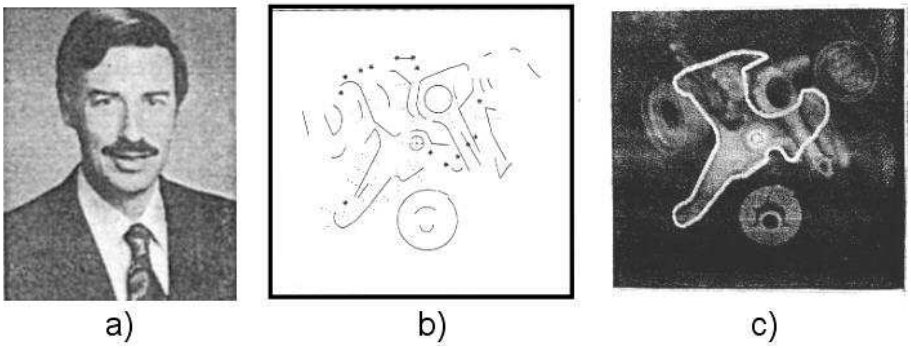


Fig. 8. Recognition of manufactured parts using a planar model. a) Walter Perkins. b) A set of point and curve features, extracted by bottom-up processing. c) The part model matched to the features in b).(From [58] with permission.)

As mentioned earlier, Goad initiated the idea that an object model could be used to plan the search for features. The plan is based on selecting features that are likely to be segmented reliably and that provide strong constraints on the projection of the model into the image. Given this plan, it is not necessary to carry out extensive feature grouping and linking in advance of the recognition stage. Instead the model constraints are imposed on the image during recognition and provide the required organization.

Perhaps the first research to carry out this approach in the implementation of a complete recognition system was David Lowe [45]. An example of his recognition system, called SCERPO [3], is shown in Figure 9. The basic approach is that a consistent interpretation of a set of image features will constrain the viewing hypotheses to a single perspective viewpoint of the model. This philosophy of minimal feature organization and strong model constraints quickly became a compelling research focus during the early half of the 1980s [10, 29, 4]. An example of recognition with essentially ungrouped features is shown in Figure 10. This work by Eric Grimson and Tomas Lozano-Perez generated considerable enthusiasm for complete reliance on prior object models for the organization of features and the detection of objects under high degrees of occlusion and shadowing. Indeed, it became kind of an academic contest to see how occluded an object could be and still achieve successful recognition.

The emphasis in the early 1980s was mainly on 2-d planar shapes or 3-d objects as imaged by 3-d range cameras [11]. This restriction reduced the number of degrees of freedom for the image projection transformation relative to

[3] Spatial Correspondence, Evidential Reasoning, and Perceptual Organization.

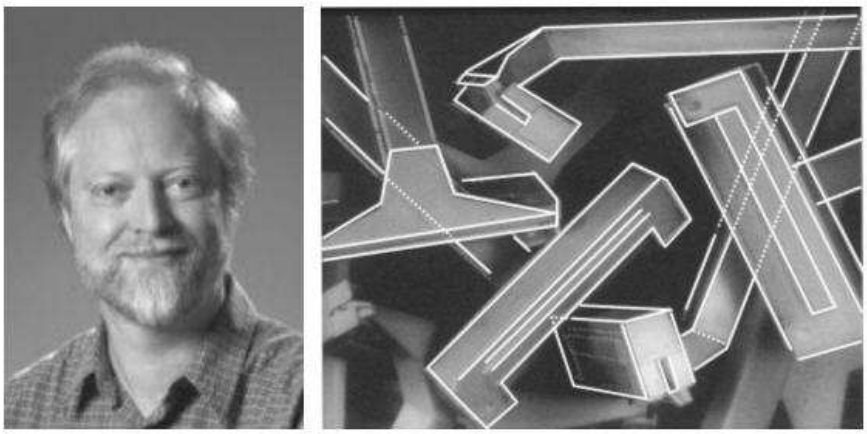


Fig. 9. Recognition based on viewpoint consistency. a) David Lowe. b)An example of recognizing plastic razors under conditions of high occlusion. (b) is taken from [42] with permission.)

the number of constraints provided by each feature-to-model assignment. There was the sense that it is important to solve 2-d planar object recognition robustly and completely before re-attacking the harder problem of 3-d object recognition from a single intensity image.

The 2-d recognition approaches were driven by a search for model-to image-transformations based on the a small number of un-grouped features. Eric Grimson exploited the *interpretation tree* that is a pre-compiled search plan for matching features. This approach is similar to the recognition plan ideas of Goad [27]. Katsu Ikeuchi and Takeo Kanade also developed an extensive recognition planning system that took into account both projected 3-d shape and self-occlusion in a tree-like plan structure [37]. Their object representation included 3-d orientation constraints based on photometric stereo and so might be called a 2.5-d representation.

Another 2-d approach of the period is based on the data indexing method of hashing on a minimum number of features,e.g., three points or lines for planar affine matching [43]. The minimum feature set is used to retrieve from a hash table the set of confirming features that would be visible and placed in the image according to the transform computed from the search features. A match is declared if the hashed features are sufficiently confirmed in the image.

It would be fair to say that the 2-d problem is now solved for many cases of practical interest such as industrial inspection and robotic placement. However, high background complexity along with expected significant occlusion can still

confound existing 2-d methods by producing a large number of false hypotheses. These recognition error statistics were studied extensively by Grimson [28].

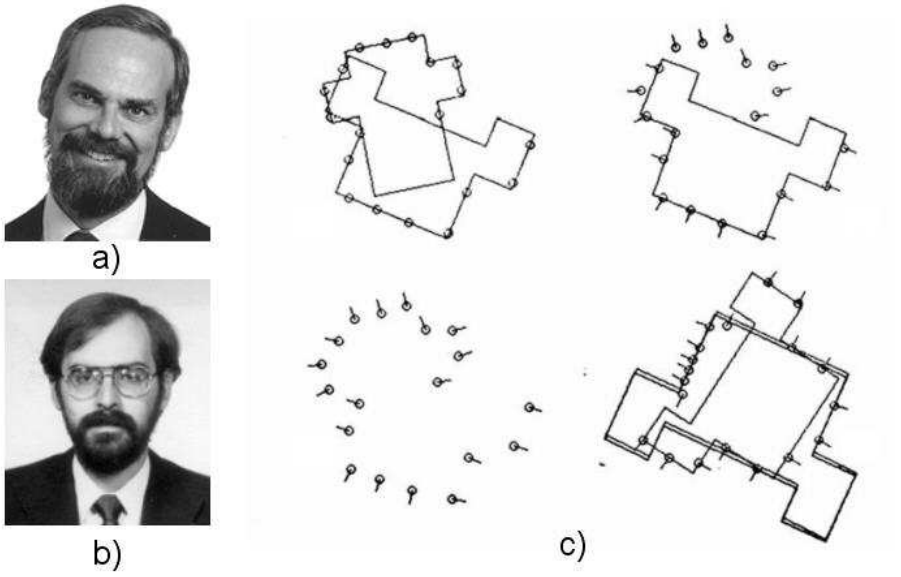


Fig. 10. The use of sparse, unorganized features for recognition. a) Eric Grimson. b) Tomas Lozano-Perez. c) Steps in forming a model recognition hypothesis based on oriented edge segments. (c) used by permission of Eric Grimson.)

By the mid 1980s, attention refocused on the recognition of 3-d objects from 2-d intensity images. These approaches exploited viewpoint consistency (equivalent to object pose consistency) where the pose was computed from a minimal set of features. The constraint of full-perspective image formation was abandoned for the use of *affine* image projection models where the camera parameters can be determined from a small number of features such as three points or a point and two intersecting lines or two lines each with a fixed point. The affine camera model, called *weak perspective* has only six parameters: tip and tilt angles, image rotation, image x-y translation and scale. Unlike full perspective camera models, the weak perspective parameters can be determined uniquely without prior camera calibration.

Again, the feature grouping problem is avoided and model hypotheses are generated directly from a match of the minimal feature set. The hypotheses can be confirmed in various ways, such as projecting the model onto the image and checking that the expected features are present (the Goad philosophy). One of the first attacks on the 3-d problem in this era was by Dan Huttenlocher and

Shimon Ullman [36]. They called the recognition process *alignment* since the image feature (in their case, a point triple) is sufficient to align the 3-d model with the image. The point triples are formed exhaustively so that the algorithm has a complexity of Mn^3, where M is the number of model triples and n is the number of feature points in the 2-d image. At the same time a similar approach

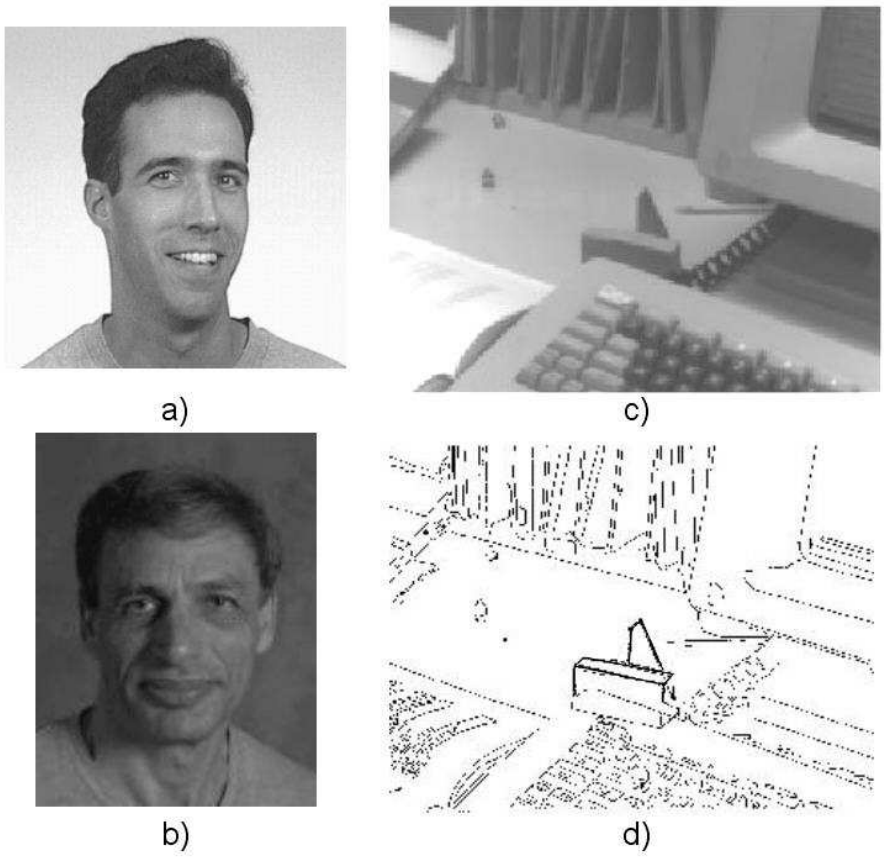


Fig. 11. Three-dimensional object recognition using alignment. a) Dan Huttenlocher. b) Shimon Ullman. c) A cluttered image. d) The aligned model, shown near the middle of the image. (c) and d) provided by Dan Huttenlocher, with permission.)

was taken by the author and Dan Thompson[78]. In their system, the model hypothesis was determined by pose clustering. The idea is that a correct object hypothesis will have all features projected into the image with the same pose. The most consistent pose is found by voting into a space of affine transformations, similar to the generalized Hough transform [5, 75]. They used a single image

feature called a vertex-pair that required that two line segments be grouped around a common vertex. Two such vertices are sufficient to determine and over-constrain the object pose. In this approach, the complexity is Mn^2, where M is the number of model vertex-pairs and n is the number of vertex pairs in the 2-d image. Reduction in matching complexity is being traded off against modest feature grouping risk. Their system was applied to the problem of aerial surveillance and achieved a respectable recognition performance for the problem of detecting aircraft at airfields with 99% accuracy. The performance result was based on extensive testing and is reported in [52].

While these viewpoint consistency approaches can overcome the lack of feature grouping, there are still limitations fundamentally caused by the absence of object features resulting from the effects itemized at the beginning of this section. The vertex-pair system, shown in Figure 12 could hallucinate the presence of models when the number of features or the tolerance on viewpoint consistency is reduced. Figure 12 d) shows numerous false positive hypotheses where support for the model is found by accident. For example the bright sidewalk region in the upper middle of the image provides strong support for the edges of the aircraft wings.

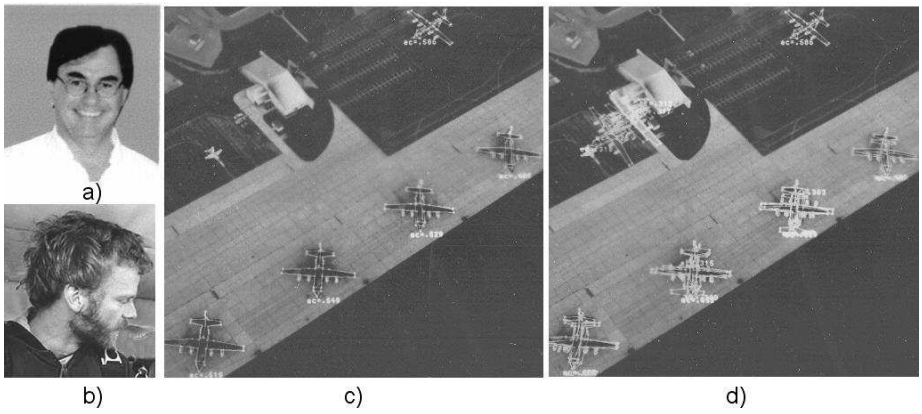


Fig. 12. The vertex-pair recognition system. a) The author. b) Dan Thompson. c) An example of aircraft recognition. d) Hallucination is possible. The same scene as c) with a relaxed tolerance to pose consistency.

These approaches based on a manually constructed 3-d object model with extra attributes to express the reliability of segmented features can be quite successful under reasonably bland backgrounds and limited amounts of occlusion. The airfield problem is particularly well-suited to these limitations. However, the approach is encumbered with the need to construct a detailed 3-d model for each specific object. In spite of this drawback, there has been extensive use of detailed

3-d models to enable target recognition. Figure 13 has thousands of polygonal surface facets and is used to recognize this specific tank in synthetic aperture radar imagery (SAR). The rationale here is that there are only a finite number of military weapons and vehicles so that a concerted effort could "model the world" in this limited domain.

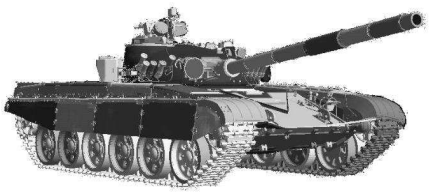


Fig. 13. A highly detailed 3-d geometric model for a tank.

6 The era of geometric invariance

By the end of the 1980s there was a rising interest in the object recognition community to move beyond the manual modeling approach and to try to automate the acquisition of models for recognition. Ideally a single view or at worst a small number of views of the object would be sufficient to construct a recognition model. A promising avenue was the concept of geometric invariance where properties of an object are determined that do not vary with viewpoint. For example under affine viewing conditions the ratio of collinear segment lengths is independent of viewpoint. That is, the length ratio in the image will be the same as in the 3-d object, regardless of affine camera parameters.

The formation of recognition models is reduced to measuring the invariant values for feature constructions that have sufficient geometric constraints to enable the formation of invariants. Objects seen under perspective are described by projective invariants such as the cross ratio and the ratio of area ratios [54]. These constructions require four collinear points and five points or five lines respectively. The configurations must not be degenerate, so that no four of the five points are collinear, for example.

The research focus was initially on planar shapes because the theory of geometric invariance for perspective and affine image formation is complete. Plane to image mappings form a transformation group and the full machinery of group invariance developed by Felix Klein and other 19th century mathematicians can be brought to bear on the recognition task. The role of projective geometry was also elevated from a minor interest, mainly relevant to the field of graphics, to a central object of study and adaptation to computer vision. Again, the results of 18th and 19th century mathematics could be readily mined for ideas to solve

the recognition task. Some of the main researchers in the geometric invariance movement are shown in Figure 14.

Fig. 14. A meeting of researchers central to the geometric invariance movement at Schenectady, New York during the month of July, 1992. Top row, left to right: Andrew Zisserman, Charles Rothwell, Luc VanGool, Joseph Mundy, Stephen Maybank and Daniel Huttenlocher. Bottom row, left to right: Thomas Binford, Richard Hartley, David Forsyth and Jon Kleinberg.

This hope of a complete theory for modeling and recognition created considerable interest in the late 1980s and early 1990s. However, the enthusiasm was tempered by two key drawbacks of representation by geometric invariance:

- it was proved independently by several researchers that no viewpoint invariants exist for general 3-d shapes [18, 14, 51];
- the grouping problem re-emerges; it is necessary to associate a rather large number of features (e.g. five lines) across views in order to check for consistent invariant values and thus a correct model hypothesis.

Nevertheless, keen interest in recognition based on invariants continued through the middle of the 1990s. It was felt that a sufficient number of classes of 3-d structures do possess invariants, such as surfaces of rotation and polyhedra, so that the lack of invariance in general does not pose a major defeat for the program. The grouping problem was sidestepped for the moment by focusing on the discovery of new invariants and integrating the representations into a complete recognition system [68, 67]. Two systems for recognition by invariants are shown in Figure 15. The recognition systems were named after characters in the Oxford-based detective stories by Colin Dexter.

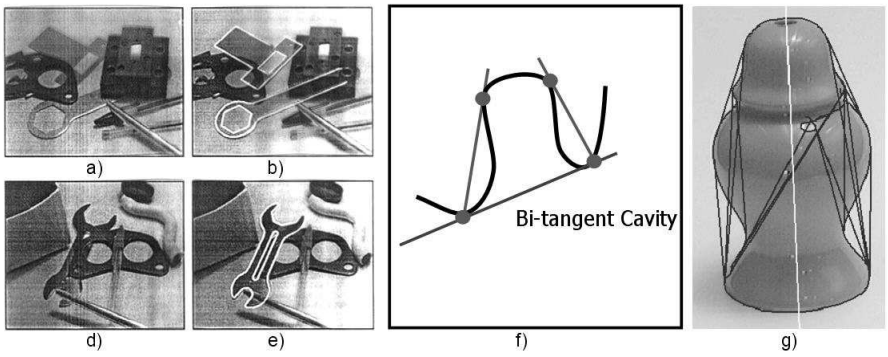


Fig. 15. Two recognition systems based on geometric invariance. a) A cluttered image with machine parts. b) Recognition of several objects by the LEWIS system using various invariant descriptions, such as five lines. c) A second image. d) Recognition by LEWIS using the invariant construction on bi-tangent cavities shown in f). Recognition of a surface of rotational symmetry by the MORSE system. The axis of rotation is recovered as well as invariants of the bi-tangent cavities.

6.1 Multiview Geometry

A complementary thread of research was intitated in 1992 by Richard Hartley and Oliver Faugueras with the goal to apply the theory of projective geometry to the relationship between multiple perspective views. An emphasis of this work was the reconstruction of 3-d geometry without the need for camera calibration. The resulting reconstruction was ambiguous up to a 3-d projective transformation and thus the central role of projective geometry in the analysis of camera configurations and reconstructed geometry.

It was quickly realized that the lack of general viewpoint invariants for a single view could be overcome if an object is seen in two or more views. Of course, one approach would be to reconstruct the 3-d geometry and then use direct 3-d recognition methods developed earlier for model-based recognition. A different approach, more in keeping with the invariance philosophy, is to derive invariants of a structure from correspondences across views. This approach is particularly attractive if the features can be easily tracked as would be the case in video image sequences. This concept was realized in recognition systems by Daphna Weinshall [82] and Stephan Carlsson [16].

From a slightly different approach one can take the position that invariants change with viewpoint but according to a set of 1-dimensional spaces. If there are sufficient constraints such as independent features on a model, it is possible to constraint the viewpoint and thus determine all the invariants for the object. In essence, the camera projection is being recovered in the invariant construction. This approach was initiated by David Jacobs [19] and extended to projective invariance by Isaac Weiss [83].

6.2 Practical issues

Feature segmentation methods had advanced little since the early 1980s [15] and the problems of missing features and noisy geometry remained. Geometric invariants are noise-prone since a minimum number of image features are used for the invariant construction. There is no redundancy to smooth out errors in feature geometry recovery. The resulting invariant values can have significant random noise variance, even within a single view [49]. In spite of these limitations, by 1995 it was possible to reliably recognize a half-dozen or so 3-d objects in somewhat cluttered scenes [86], by exploiting class-based invariance such as of surfaces of revolution and canal surfaces. However, there was the growing realization that recognition performance was not going to significantly improve. Progress would depend on better image segmentation methods, not on extensions of the lexicon of invariant structures.

In retrospect, given recent advances in video feature tracking, it would have been a much better strategy for planar object recognition to compute the plane-to-plane projective transformation using all the features in a consistent statistical optimization strategy such as RANSAC [12, 26]. With the transform known, all feature coordinates and parameters become, in effect, invariants. This same strategy could be employed for 3-d invariant calculations using mutual pose constraints among objects. This approach was not taken at the time since it was considered bad form for an invariance researcher to want to know anything about the transform parameters

7 The rise of appearance methods

At the same time as the geometric invariance program was reaching the end of its active period, new recognition approaches strongly rooted in intensity appearance were discovered: appearance manifolds [55] and affine invariant intensity features[71]. Shree Nayar's system was based on SLAM [4] which is a C library of tools for processing images taken over a large number of viewpoints and lighting conditions. The input image set is compiled into a continuous eigen-space of the image intensity covariance, treating the entire image as a 1-d vector.

Recognition is achieved by finding the appearance space *closest* to the input image. In SLAM, distance is computed as Euclidean distance on a low-dimensional subspace representing the largest eigenvalues. The SLAM algorithm produced very impressive results with high recognition rates on a large library of objects. Remarkably, no model assumptions or image segmentation is required and the recognition hypothesis carries with it an estimate of the object's 3-d pose. Nayar's work generated tremendous interest, overshadowing ongoing recognition research based on geometry. There was renewed interest in understanding intensity appearance phenomena [6] and in the development of invariance to illumination changes [72].

[4] Software Library for Appearance Modeling

The geometry recognition community remained somewhat skeptical of the power of global appearance methods, such as SLAM, particularly with respect to the ability to withstand occlusion. In conjunction with a representation workshop in 1996 it was decided to carry out a comparison between SLAM and MORSE [53]. The experiments focused on surfaces of revolution (SOR). A set of images of SORs at different tilt angles was collected under varying degrees of occlusion. Recognition by SLAM was carried out using the standard nearest point algorithm while recognition in MORSE was based on invariants of the bi-tangent cavities formed on the outline of the SOR. The appearance manifold for example SORs and the MORSE results are shown in Figure 16. The result

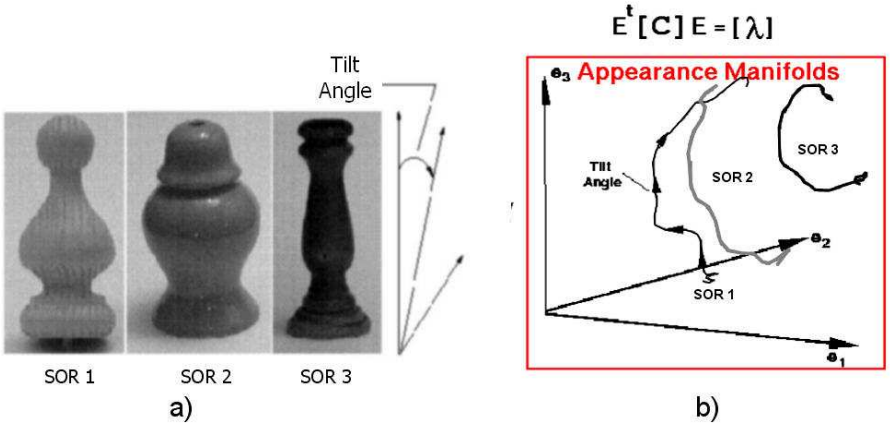


Fig. 16. SLAM vs MORSE. a)Example surfaces of revolution from the experiment. b) The SLAM appearance manifolds for the SORs.

of the comparison was very surprising – there was no clear winner. The presence of limited amounts of occlusion could be handled by SLAM as well as MORSE. Both systems faired badly under heavy occlusion. It is not well-understood why the global appearance manifold is somewhat immune to occlusion. Perhaps eliminating the higher order eigenvectors smears out the perturbations of occlusion so that the final manifold distance value is not much affected. In any case, the ability of SLAM to learn an effective 3-d recognition model for any object fully automatically without any explicit geometric representation was a compelling paradigm that set the stage for recognition research over the next decade.

The problem of occlusion in appearance methods can be solved by using more local intensity features such as planar regions about interest points. The successful application of this idea by Cordelia Schmid and Roger Mohr [72] inspired an intensive search for other intensity and affine projection invariant features [46, 70, 79, 38, 50]. The basic assumption is that intensity regions are derived from

locally planar surface patches and viewed by an affine camera. Thus, local affine constructions such as ratios of areas can be used to determine consistent feature matches. A more global 3-d viewpoint consistency constraint can be invoked by deriving the fundamental matrix from hypothesized matches. Any correct match would be consistent with the epipolar geometry of the two views [32]. The recognition strategy is to generate hundreds of affine patch features and then sift them into object hypotheses by geometric match consistency.

In this approach object models are learned directly from a set of images without geometric segmentation, except for the detection of local corners or other interest operators. The models can be acquired at the video frame rate and recognition can also be carried out in real time [5]

Another impressive achievement using affine patches is the Video Google system by Josef Sivic and Andrew Zisserman [73]. Affine patch features are derived and their geometric relations pre-compiled for each frame of a feature length film (100,000 frames). This preprocessing step is similar to Goad's strategy, described in Section 4, to divert expensive combinatorial operation to an off-line compilation process. After compilation process, an object can be designated in one frame and matches found in any other frame of the movie in seconds by exploiting the pre-compiled relations between the extracted features.

More recently, the affine patch features have been integrated into a 3-d representation [66]. A 3-d model is constructed from a set of affine patches arranged to tessellate the surface of the object. The patch arrangement is derived from a dense set of multiple views of the object. Instead of purely geometric features such as the polygonal facets used by Roberts, a 3-d object is represented by features that are easy to find over a wide range of camera viewpoints. Full feature coverage over the viewsphere is obtained by a combination of manual selection and automated feature refinement. Issues such as self-occlusion are handled naturally by the 3-d structure as has always been the case for purely geometric methods. The constraint of viewpoint consistency is also exploited during the recognition process to rule out false matches.

Affine patches have also been exploited as *parts* in a new attack on the problem of generic object recognition [24, 44]. The rationale is that invariant regions provide a stable description of objects and that a degree of flexibility in the geometric relationships between patches can account for in-class variations. One is guaranteed that parts defined in this way can be reliably segmented, an essential requirement for generic object recognition.

8 Coming full circle?

One way to look at the current state of object recognition research is that the four decade dependence on step edge detection for the construction of object features has been broken. Step edge boundaries are still useful in forming an object description where the object surface is bland and free of surface markings.

[5] The author viewed an impressive live demonstration of the SIFT recognition system by David Lowe in 2003 [61]

But, for a large fraction of object surfaces and textures, affine patch features can be reliably detected without having to confront the difficult perceptual grouping problems that are required to form purely geometric boundary descriptions from edges.

Some revisiting of the earlier themes of geometry-based object recognition can be expected as the affine patch feature vocabulary is woven into the edge-based prior art. For example, one can envision affine-patch aspect graphs where the aspect cells are based on continuous measures of the variability of the affine properties of a patch. In this case, the cell boundary represents the removal and insertion of patches required to maintain good recognition performance. The problem of aspect scale is mitigated since the patch segmentation automatically adapts to the granularity of visible features [6]

The use of viewpoint consistency has been an integral part of the geometric recognition strategy since the beginning and is essential in filtering match hypotheses. General 3-d relations among patches are enforced by the epipolar constraint and local planarity relations can be tested by affine invariant relations among patches. However, if patches are treated as isolated features, it quickly becomes combinatorially impractical to rely on large degree n-ary patch relations to constrain match integrity. This combinatorial problem can be solved by re-introducing the classic role of generic shape models such as polyhedra and generalized cylinders.

The constraints that must exist between faces for a connected polyhedral surface [76] can be exploited to confirm feature matches and at the same time define the 3-d polyhedral shape [7]. A similar idea could be applied to generalized cylinder parts where the local "flow" of individual patch-to-image transforms can define the axis and boundaries of the cylinders. This extended representation can bridge the gap between the relatively local, but reliably detected, affine regions and more meaningful GC object components (parts) that are difficult to segment from step edge boundary information alone.

Global shape recovery from local estimates of affine properties was exploited by Jan Koenderink in his study of the capability of the human visual system to estimate surfaces from local orientation [40]. In this work, local surface normals were integrated to form a 3-d surface. The combination of local orientations from affine patches could also be used to enable the recovery of surface geometry as a first step to recover generic shape descriptions.

In summary, it is certain that the role of geometric representations of objects in recognition will not be displaced for long. Beyond mere statistical dependence,there seem to be only two avenues to a theory of object class: geometry

[6] This kind of aspect graph was implemented for the vertex-pair matcher, based on the expected variance in the affine transformation computed from a given model vertex-pair as a function of viewpoint [52]. Also, the system by Art Pope and David Lowe [63] used a kind of aspect graph based on the probability of feature detection with respect to viewpoint.

[7] The polyhedral faces must have at least four sides to generate constraints, but for complex enough shapes, patch arrangements can be designed to satisfy Sugihara's constraint system.

and function. Moreover, the characterization of function is itself largely couched in geometry along with the laws of physics [74]. Such models are essential to fuse statistical class correlations across scene contexts and to arrive at a formal understanding of categories. To quote Larry Roberts from four decades ago, 'The perception of solid objects is a process which can be based on the properties of three-dimensional transformations and the laws of nature.'

Acknowledgments

The author is honored to have been part of the geometric era and to have met and worked with many of the researchers that remain committed to understanding the mysteries of the recognition task. The author is particularly indebted to Thomas O. Binford for his thoughtful and determined effort to enlighten and inspire.

References

1. G. Agin and T. Binford. Computer description of curved objects. In *Proceedings 3rd International Conference on Artificial Intelligence*, pages 629–640, 1993.
2. G. J. Agin. *Representation and Description of Curved Objects*. PhD thesis, Stanford University, October 1972.
3. A. Ambler, H. Barrow, C. Brown, R. Burstall, and R. Popplestone. A Versatile Computer-Controlled Assembly System. In *International Joint Conference on Artificial Intelligence*, pages 298–307, 1973.
4. N. Ayache and O. Faugeras. HYPER: A New Approach for the Recognition and Positioning of Two-Dimensional Objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(1):44–54, January 1986.
5. D. Ballard. Generalizing the Hough Transform to Detect Arbitrary Shapes. *Pattern Recognition*, 13(2):111–122, 1981.
6. P. Belhumeur and D. Kriegman. Learning and recognizing objects using illumination subspaces. In *Proceedings of the IEEEConference on Computer Vision and Pattern Recognition*, pages 270–277, 1996.
7. I. Biederman. Human Image Understanding: Recent Research and a Theory. *Computer Vision, Graphics and Image Processing*, 32:29–73, 1985.
8. T. O. Binford. Visual Perception by Computer. *Proc. IEEE Conf. on Systems and Control*, December 1971.
9. T. O. Binford. Spatial understanding: the successor system. In *Proceedings of the ARPA Image Understanding Workshop*, pages 12–20. Defense Advanced Research Projects Agency, Morgan Kaufmann Publishers, Inc., 1992.
10. R. Bolles and R. Cain. Recognizing and locating partially visible objects: The local-feature-focus method. *International Journal of Robotics Research*, 1(3):57–82, 1982.
11. R. Bolles and R. Horaud. 3DPO: A Tree-dimensional Part Orientation System. *International Journal of Robotics Research*, 5(3):3–26, 1986.
12. R. C. Bolles and M. A. Fischler. A RANSAC-based approach to model fitting and its application to finding cylinders in range data. In *International Joint Conference on Artificial Intelligence*, pages 637–643, Vancouver, Canada, August 1981.

13. R. Brooks. Symbolic reasoning among 3D models and 2D images. *Artificial Intelligence Journal*, 17:285–348, 1982.
14. J. Burns, R. Weiss, and E. Riseman. *The Non-existence of General-case View-Invariants*, pages 120–131. MIT Press, 1992.
15. J. F. Canny. Finding edges and lines in images. Technical Report AI-TR-720, Massachusets Institute of Technology, Artificial Intelligence Laboratory, June 1983.
16. S. Carlsson. Multiple image invariance using the double algebra. In J. L. Mundy, A. Zissermann, and D. Forsyth, editors, *Applications of Invariance in Computer Vision*, volume 825 of *Lecture Notes in Computer Science*, pages 145–164. Springer-Verlag, 1994.
17. I. Chakravarty. The use of characteristic views as a basis for the recognition of three-dimensional objects. *Proc. Society for Photo-Optical Instrumentation Engineers conference on Robot Vision*, 336:37–45, May 1982.
18. D. Clemens and D. Jacobs. Space and time bounds on model indexing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(10):1007–116, 1991.
19. D. T. Clemens and D. W. Jacobs. Model group indexing for recognition. In *Proceedings of the IEEEConference on Computer Vision and Pattern Recognition*, pages 4–9, Maui, HI, June 1991.
20. M. B. Clowes. On seeing things. *Artificial Intelligence Journal*, 2:79–116, 1971.
21. C. Cyr and B. Kimia. 3d object recognition using shape similiarity-based aspect graph. In *Proceedings of the International Conference on Computer Vision*, pages 254–261, Vancouver, Canada, July 2001.
22. S. Dickinson, A. 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*, 14(2):174–198, 1992.
23. O. Faugeras, J. Mundy, N. Ahuja, C. Dyer, A. Pentland, R. Jain, K. Ikeuchi, and Bowyer K. Why aspect graphs are not (yet) practical for computer vision. In *IEEE Workshop on Directions in Automated CAD-Based Vision*, pages 98–104, 1991.
24. R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 264–271, June 2003.
25. O. Firschein, editor. *RADIUS: Image Understanding for Imagery Intelligence*. Morgan Kaufmann, San Francisco, 1997.
26. A. W. Fitzgibbon and A. Zisserman. Automatic 3D model acquisition and generation of new images from video sequences. In *Proceedings of European Signal Processing Conference (EUSIPCO '98), Rhodes, Greece*, pages 1261–1269, 1998.
27. C. Goad. Special purpose automatic programming for 3d model-based vision. In *Proc. DARPA Image Understanding Workshop*, pages 94–104, Arlington, VA, June 1983.
28. W. E. L. Grimson. *Object Recognition by Computer: The Role of Geometric Constraints*. The MIT Press, Cambridge, Massachusetts, London, England, 1990.
29. W. E. L. Grimson and T. Lozano-Pérez. Model-based recognition and localization from sparse range or tactile data. *International Journal of Robotics Research*, 3(3):3–35, 1984.
30. A. Guzman. Decomposition of a visual scene into three-dimensional bodies. In *Proceedings Fall Joint Computer Conference*, volume 33, pages 291–304, 1968.
31. A. Guzman. Analysis of curved line drawings using context and global information. In B. Meltzer and D. Michie, editors, *Machine Intelligence 6*, pages 325–375. John Wiley and Sons, Inc., New York, NY, 1971.
32. R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521623049, 2000.

33. B. K. P. Horn. Shape from shading: a method for obtaining the shape of a smooth opaque object from one view. Technical Report TR-79, MIT Project Mac, October 1970.
34. M. Hu. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, 8(2):179–187, February 1962.
35. D. A. Huffman. Impossible Objects as Nonsense Sentences. In B. Meltzer and D. Michie, editors, *Machine Intelligence 6*, pages 295–324. Edinburgh University Press, 1971.
36. D. P. Huttenlocher and S. Ullman. Object recognition using alignment. In *Proceedings of the First International Conference on Computer Vision, London*, pages 102–111, 1987.
37. K. Ikeuchi and T. Kanade. Applying sensor models to automatic generation of object recognition programs. In *Proc. Second Int'l Conf. Comput. Vision*, pages 228–237, Tampa, FL, December 1988.
38. T. Kadir, A. Zisserman, and M. Brady. An affine invariant salient region detector. In *Proceedings of the 8th European Conference on Computer Vision, Prague, Czech Republic*, May 2004.
39. J. J. Koenderink and A. J. van Doorn. The singularities of the visual mapping. *Biological Cybernetics*, 24:51–59, 1976.
40. J. J. Koenderink and Andrea J. van Doorn. Relief: pictorial and otherwise. *Image and Vision Computing.*, 13(5):321–334, 1995.
41. D. Kriegman and J. Ponce. Computing exact aspect graphs of curved objects:solids of revolution. *The International Journal of Computer Vision*, 5(2):119–136, November 1990.
42. R. Kurzweil. *The age of intelligent machines*. MIT Press, Cambridge, MA, 1990.
43. Y. Lamdan and H.J. Wolfson. Geometric Hashing: A General and Efficient Model-Based Recognition Scheme. In *Proceedings of the 2nd International Conference on Computer Vision, Tampa, Florida*, pages 238–249, December 1988.
44. S. Lazebnik, C. Schmid, and J. Ponce. Semi-local affine parts for object recognition. In *British Machine Vision Conference*, volume volume 2, pages 779–788, 2004.
45. D. Lowe. *Perceptual Organization and Visual Recognition*. Kluwer Academic Publishers, 1985.
46. D. G. Lowe. Object recognition from local scale-invariant features. In *ICCV '99: Proceedings of the International Conference on Computer Vision-Volume 2*, page 1150, Washington, DC, USA, 1999. IEEE Computer Society.
47. A. K. Mackworth. Interpreting pictures of polyhedral scenes. *Artificial Intelligence Journal*, 4:99–118, 1973.
48. D. Marr. *Vision*. W.H. Freeman and Co., 1982.
49. P. Meer, S. Ramakrishna, and R. Lenz. Correspondance of coplanar features through p^2-invariant representations. In J. L. Mundy, A. Zissermann, and D. Forsyth, editors, *Applications of Invariance in Computer Vision*, volume 825 of *Lecture Notes in Computer Science*, pages 437–492. Springer-Verlag, 1994.
50. K. Mikolajczyk, T. Tuytelaars, C. Schmid, J. Zisserman, A.and Matas, F. Schaffalitzky, T. Kadir, and Van Gool L. A comparison of affine region detectors. *Int. J. Comput. Vision*, To Appear, 1994.
51. Y. Moses and S. Ullman. Limitations of non model-based recognition systems. In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision*, volume 588, pages 820–828, Santa Margherita Ligure, Italy, May 1992. Springer-Verlag.

52. J. L. Mundy and A. J. Heller. The evolution and testing of a model-based object recognition system. In *Proceedings of the 3rd International Conference on Computer Vision*, pages 268–282, Osaka, Japan, December 1990. IEEE Computer Society Press.
53. J. L. Mundy, A. Liu, N. Pillow, A. Zisserman, S. Abdallah, S. Utcke, S. K. Nayar, and C. Rothwell. An experimental comparison of appearance and geometric model based recognition. In *Object Representation in Computer Vision*, pages 247–269, 1996.
54. J. L. Mundy and A. Zisserman, editors. *Geometric Invariance in Computer Vision*. MIT Press, 1992.
55. H. Murase and S. Nayar. Learning and recognition of 3d objects from appearance. *The International Journal of Computer Vision*, 14(1):5–24, 1995.
56. R. Nevatia and T. O. Binford. Structured descriptions of complex obects. *Proc. 3rd International Joint Conference on Artificial Intelligence*, pages 641–647, 1973.
57. R. Nevatia and T. O. Binford. Description and Recognition of Curved Objects. *Artificial Intelligence Journal*, 8:77–98, 1977.
58. W. Perkins. A model-based vision system for industrial parts. *IEEE Transactions on Computers*, C-27(2):126–143, February 1978.
59. S. Petitjean. The complexity and enumerative geometry of aspect graphs of smooth surfaces. April 1994.
60. H. Plantinga and C. Dyer. Visibility, occlusion and the aspect graph. *The International Journal of Computer Vision*, 5(2):137–160, November 1990.
61. J. Ponce. Designing tomorrow's category-level 3D object recognition systems: an international workshop. Taormina, Sicily, September 2003.
62. J. Ponce, A. Zisserman, and M. Hebert, editors. *Object Represenation in Computer Vision II*, volume 1144 of *Lecture Notes in Computer Science*, Cambridge, UK, June 1996. Springer-Verlag.
63. A. Pope and D. Lowe. Learning Appearance Models for Object Recognition. In Ponce et al. [62], pages 201–219.
64. L. G. Roberts. Machine perception of three-dimensional solids. In Tippett, J. and Berkowitz, D. and Clapp, L. and Koester, C. and Vanderburgh, A., editor, *Optical and Electrooptical Information processing*, pages 159–197. MIT Press, 1965.
65. A. Roland and P. Shiman. *DARPA and the Quest for Machine Intelligence*. MIT Press, Cambridge, 2002.
66. F. Rothganger, S. Lazebnik, C. Schmid, and J. Ponce. 3d object modeling and recognition using affine-invariant patches and multi-view spatial constraints. In *CVPR*, pages 272–280, 2003.
67. C. Rothwell. *Object recognition through invariant indexing*. Oxford University Science Publications. Oxford University Press, February 1995.
68. C. A. Rothwell, D. A. Forsyth, A. Zisserman, and J.L. Mundy. Extracting projective structure from single perspective views of 3D point sets. In *Proceedings International Joint Conference on Computer Vision*, pages 573–582, Berlin, Germany, May 1993. IEEE Computer Society Press.
69. S. Sarkar and K. L. Boyer. Perceptual organization in computer vision: A review and a proposal for a classificatory structure. *IEEE Transactions on Systems, Man, and Cybernetics*, 23:382–399, 1993.
70. F. Schaffalitzky and A. Zisserman. Multi-view matching for unordered image sets, or "How do I organize my holiday snaps?". In *Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark*, volume 1, pages 414–431, 2002.

71. C. Schmid, P. Bobet, B. Lamiroy, and R. Mohr. An image-oriented cad approach. In Ponce et al. [62], pages 221–246.
72. C. Schmid and R. Mohr. Local greyvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5):530–535, 1997.
73. J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. In *Proceedings of the International Conference on Computer Vision*, October 2003.
74. L. Stark and K. Bowyer. Generalized Object Recognition through Reasoning About Association of Function to Structure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:1097–1104, 1991.
75. G. Stockman. Object recognition and localization via pose clustering. *Computer Vision, Graphics, and Image Processing*, 40:361–387, 1987.
76. K. Sugihara. *Machine Interpretation of Line Drawings*. MIT Press, 1986.
77. M. J. Tarr and S. Pinker. When does human object recognition use a viewer-centered reference frame? *Psychological Science*, 1(42):253–256, 1990.
78. D. W. Thompson and J. L. Mundy. Three-dimensional model matching from an unconstrained viewpoint. In *Proceedings of the International Conference on Robotics and Automation, Raleigh, NC*, pages 208–220, 1987.
79. T. Tuytelaars and L. Van Gool. Matching widely separated views based on affine invariant regions. *Int. J. Comput. Vision*, 59(1):61–85, 2004.
80. S. A. Underwood and C. L. Coates. Visual Learning from Multiple Views. *IEEE Transactions on Computers*, C-24(6):651–661, 1975.
81. D. Waltz. Understanding line drawings of scenes with shadows. In Patrick H. Winston, editor, *The Psychology of Computer Vision*, pages 19–91. McGraw-Hill, 1975.
82. D. Weinshall and C. Tomasi. Linear and incremental acquisition of invariant shape models from image sequences. In *Proceedings International Joint Conference on Computer Vision*, pages 675–682, Berlin, Germany, 1993. IEEE Computer Society Press.
83. I. Weiss and M. Ray. Model-based recognition of 3d objects from single images. *PAMI*, 23(2):116–128, February 2001.
84. P. H. Winston. The MIT robot. In B. Meltzer and D. Michie, editors, *Machine Intelligence 7*, pages 431–463. Edinberg University Press, 1972.
85. M. Zerroug and R. Nevatia. From an intensity image to 3-d segmented descriptions. In J. Ponce, M. Hebert, and A. Zisserman, editors, *Object Representation in Computer Vision II*, pages 11–24, 1996.
86. A. Zisserman, J. Mundy, D. Forsyth, J. Liu, N. Pillow, C. Rothwell, and S. Utcke. Class-based grouping in perspective images. In *Proceedings of the 5th International Conference on Computer Vision*, pages 183–188, Boston, MA, June 1995. IEEE Computer Society Press.