IV 3D MODEL-BASED VISION PROJECT

Introduction by the Editors

A THE GRANT PROPOSAL (Written 1983)

1 OBJECTIVES

The overall objective of the 3D Model-Based Vision Project is to develop a scheme for the recognition and manipulation of 3D objects using information about the 3D structure of their visible surfaces delivered by the 2.5D Sketch Project.

Object recognition has two major components. First, there must be a collection of stored model descriptions cast within some representational scheme; and second, there must be one or more ways of associating descriptions derived from images with descriptions in the collection of models (Marr and Nishihara, 1978). The 3D Model-Based Vision Project reflects these twin requirements by having two separate, albeit closely inter-related, sub-projects. The first has as its goal the design of a representational scheme for 3D objects which is specifically tailored to address the issues and problems of describing visual objects using the kind of data structures carried by the 2.5D Sketch. This is called the YASA project. The second will be concerned with developing methods for accessing the collection of stored object models from a newly derived 2.5D Sketch. This is called the 3D Model Invocation and Verification Project.

The 3D Model-Based Vision Project will be led by Dr J E W Mayhew, working in conjunction with Dr R J Popplestone[1] in Edinburgh (Department of Artificial Intelligence) who wishes to interface 3D vision capabilities to the robot programming language RAPT-2.

2 THE YASA PROJECT

The primitives to be used in the YASA[2] recognition scheme will be 3D surface features and relationships between them. The 2.5D Sketch data structures to be used for input for YASA can be regarded as equivalent to what Brooks (1981) called the 'observation graph' but with an important difference. In the scheme proposed here both the primitives and the relationships expressed between them in the graph are three dimensional.

Brooks points out that the most important factor in predicting shape is the orientation of the object relative to the camera. The two factors which determine the 2D image are: (i) the self-occlusion relationships of the object given a specific camera geometry (ignoring other object occlusions!), and (ii) the metrical distortions produced by the projection from 3D onto 2D. In the 3D depth map the self-occlusion relationships are almost the same as for the 2D case and the representation we have chosen for the description of 3D objects (based on what Koenderink and van Doorn (1979) have called the 'visual potential' of the object) explicitly recognises and exploits self occlusion relationships. However, the depth map is not subject to the same metrical distortions as the 2D image, for whereas the 2D projection of a right angle may be acute or obtuse depending on its orientation to the imaging plane, a right angle projects into the depth map as a right angle whatever its orientation (within obvious noise and resolution limits). Thus a very important characteristic of the scheme we propose is that the matching is between 3D structures extracted from the depth map and 3D structures in the object model catalogue, and not as is the case in many current object recognition schemes, between 2D structures predicted from the object model and imaging geometry. The proposed scheme has much in common with the 3DPO project described by Bolles, Howard and Hannah (1983).

It is realised of course that image aquisition limitations may degrade the 3D data and an important concern of the YASA scheme will be the development of methods that degrade gracefully.

Binford (1982) lists several criteria for a representation of 3D shape:

a) The scope of the representation should be such that a wide range of objects can be classified and described with it.

b) The primitives of the scheme should be locally computable. They should not be simple volumetric prototypes but, rather like splines, be both general and locally generated from the imaged data.

c) The representation should make the similarity relationships between object parts and between object wholes easy to recognise and describe.

d) The appearance of the objects should be readily predicted.

e) There should be a natural coarse to fine segmentation of the object into whole/part relationships. The parts should be volumes and locally realisable, with a complex object being the 'glued' union of simpler components.

There are few differences between these and the following criteria proposed by Marr and Nishihara (1978):- primitives should be accessible (readily computed from the image data); the scope should be large and descriptions canonical (unique); and both similarities and differences between shapes should be describable.

Both Binford's and Marr and Nishihara's criteria are based on an important assumption, namely that the task domain is one of visible object recognition. The criteria that govern the design of CAD/CAM body modelling schemes are based on different requirements but nevertheless various CAD/CAM techniques have helped shape the present proposal.

---

[1] As noted previously, Robin Popplestone left Edinburgh soon after the consortium began its work, with first Pat Fothergill and then Bob Fisher taking over his role.

[2] YASA: *Y*et *A*nother *S*illy *A*cronym. Sorry - JEWM.

Recent developments in constructive solid geometry (CGS) schemes (eg PADL - see Requicha, 1980, Brown, 1982) are attempts to overcome the representational deficiencies of earlier wire frame and boundary representation CAD/CAM systems. The major problems associated with the early schemes was the difficulty in checking the 'objects' for inconsistencies, such as hanging edges or faces, and hence for automating the computations of mass properties of the object as well as its graphical display. In CSG schemes solids are represented as a tree of ordered additions and subtractions of simple volumetric primitives (eg cylinders, cubes, cones, spheres etc, the bounded intersections of quadric halfspaces) using a regularised set of Boolean operators and rigid motions. (Regularised operators prevent the construction of objects with hanging faces and edges and thus help ensure the geometrical validity of the objects). Though a considerable advance over earlier wireframe systems, in order for a designer to see what he has 'built' using CSG it is necessary for the representation to be converted to a boundary representation before it can be input to a computer graphics display.

Following discussion with C Brown at Rochester (PADL) and members of the Leeds Body Modelling Project, it is clear that notwithstanding its considerable advantages for design, CSG is not a suitable scheme for the purposes of object recognition (in terms of Binford's criteria it satisfies only criterion (a): its scope is estimated as 95% of manufactured parts; Requicha and Voelcker (1982). Moreover, if as seems likely the automation of manufacture will increasingly utilise CSG schemes then it seems that another criterion to be added to those given by Binford is that there be a valid and reliable conversation from the 3D shape representation used for body modelling and the 3D shape representation used for visual recognition and validation. The calculation of the boundary representation from the CSG representation is an example of an exact conversion between representations.

We have chosen as the primitives of representation entities which are both importantly related to the 3D geometry of the object and also readily identified or inferred from their projections into the depth map, such as the qualitative and quantitative 3D descriptions of vertices, edges and surface regions, their 3D relationships, and some simple global gestalts (see the 2.5D Sketch Project). A possible choice for the organisation of these primitives in an object model that would facilitate model invocation and verification is a data structure somewhat similar to the winged edge representation proposed by Baumgart (1972) but one which takes directly into account the viewing geometry. Here we can exploit hidden surface removal priority sorting ideas derived from computer graphics; and also what have variously been called invariant and quasi-invariant features (Attneave, 1954; Brooks, 1981), characteristic views (Hrechanyk and Ballard, 1982, and many others!), and the visual potential of the object (Koenderink and van Doorn, 1979).

The proposed representation for YASA is a hierarchical organisation of clusters of 3D features which are stable over variations in viewpoint. Consider a 3D object centred in a transparent sphere of relatively large radius compared to the depth variation in the object. Let an eye wander over the surface of the sphere, marking on it the boundaries of regions within which specific 3D surface features of the object can be seen. Repeating this operation for all the features will produce a map of the object's viewing potential containing regions of different sizes, inclusion/exclusion relationships, and degrees of overlap, with the whole defining a hierarchical organisation of clusters of 3D features, their transformations and occlusion relationships.

Another way of illustrating the proposed representation is in terms of volumes produced by the intersections of the complements of the half spaces of the boundary surfaces. Consider a planar boundary surface of an (opaque) object. It can only be seen if the viewpoint is on the non-object side of the surface. If three planes intersect to form a corner then they define a quadrant in space in which the three faces of the corner are potentially visible. That is, the junction, the edges, and the faces form a stable 3D feature cluster which is visible until the viewpoint moves out of the quadrant, the latter being an example of what will hereafter be termed a 'view solid'. The union and intersection of view solids produced by other feature clusters provides an organisation of the view potential and suggests that procedures similar to those used in the construction of the 3D solid may provide a starting point for development of methods for the computation of the object description.

There are always problems arising from occlusion, particularly those arising from the boundaries of smooth surfaces. A smooth surface can give rise to an image feature and discontinuity in the depth map as the line of sight becomes tangent to the surface. Such a feature is called an 'extremal boundary' and its status in solid body modelling boundary representations is recognised as problematic because, unlike the edges arising from the junctions of surfaces, the position of the extremal boundary is viewpoint dependent and therefore not easily represented in object centred coordinate system. For objects of revolution or rotation the extremal boundary is an important shape descriptor and the 3D space curve corresponding to the swept function is trivially part of the representation proposed here (it will be associated with a very large view solid). Another issue is occlusion discontinuities. From some viewpoints a surface will project an extremal boundary, but from others, like the nose on a face, it will not. It is envisaged that the YASA scheme will be capable of representing this sort of information, if only in qualitative and heuristic fashion.

In terms of Binford's list of criteria, the YASA representation satisfies, at least partially, all except possibly (e), i.e. that there should be a natural segmentation of the object into part/whole relationships in which the parts are locally realisable volumes. In this regard it is possible that in some cases the hierarchical nature of the YASA representation is such that a particular cutting plane would segment the object into two components, though the reason for wanting to do this may not be obvious. Possibly of greater potential application is the operation in the opposite direction, i.e. in the construction of an object out of component parts. It may be necessary to backup to the level of the CSG representation to recompute the boundary representation and from that compute the stable feature clusters that comprise the viewing potential of the new object but whether a method that merges viewing potential can be developed will need to be investigated. This issue is of particular relevance in assembly task applications that use visual verification.

3 SUMMARY OF YASA REPRESENTATION

The basis of the YASA scheme is a form of winged edge surface representation describing the 3D geometrical relations of surface features in an object centred coordinate system. It is equivalent to the boundary representation of the body

model in so far as all the vertices, edges, and faces comprising the visible surface of the object are explicitly identified and described and the boundary representation could be generated for display if required (although the representation will be somewhat richer than the boundary representation as meta-feature gestalts or groupings will also be included as part of the object's description).

If there is any novelty in the YASA scheme it is by virtue of the proposal to group subsets of the 3D feature-nodes of the winged edge graph on the basis of their stability over variations in viewpoint. Thus, if the winged edge graph implicitly describes the complete viewing potential of the object the hierachical organisation of the 3D feature-node clusters is an explicit description of the viewing potential of the object that is invariant over a particular range of viewpoints but may change catastrophically outside that range. Included in this organisation of the graph will be information concerning:- any meta-feature or gestalt descriptions of the particular stable feature cluster; feature transformations (eg a edge may project as an orientation discontinuity over a certain range of viewpoints and as a depth discontinuity afterwards); and possible potential extremal boundary/occlusion relationships of smooth curved surfaces.

B WHAT REALLY HAPPENED?

The YASA Project began with Mayhew writing a series of internal AIVRU memos and a simple CSG body modeller and ray caster to explore their implications. These were written while Mike Gray of IBM Winchester wrote a boundary file evaluator for the IBM body modeller WINSOM and extended it to make explicit the external surface intersections (the surfaces, edges and vertices) organised on the basis of their visibility from viewpoints around a tesselation of the sphere surrounding the object. Gray's work was reported to an Alvey Vision Conference but regrettably he left the project after about 18 months and before producing a formal paper, which is why no report of his work is included here (Gray's paper did have some influence on the psychophysical and neurophysiological work of Perrett: see e.g. Perrett and Harries, 1988). Gray's replacement at IBM also left the project before a publishable report was attained. Thus the proposed extension of the representation to include extremal boundaries and virtual vertices (eg edges and vertices that occur only in the 2D projections of occlusion relationships between surfaces) was not completed.

Following John Knapman's assignment to leadership of the IBM effort in the consortium a change of direction in their work was made towards the Wireframe Completion Project (see Knapman's paper on cyclide patches [22]). Knapman also produced a working demonstration of a system for 3D polygonal model identification from stereo data, though not one of the envisaged YASA type [21].

As no personnel were available in AIVRU to pursue in detail the YASA-related ideas developed in Mayhew's memos, the project lapsed, though if one looks hard, traces of the kind of thinking it engendered can be found in Fisher's SMS [24].

Since the proposal was written several object recognition schemes of a very similar kind to that proposed for the YASA Project have been published. Generously interpreted, these suggest that the fundamental ideas on which it was based (which can be traced to those of Koenderink) were well-founded. An extension of Koenderink's work has been published by Joachim Rieger of GEC (Rieger, 1987, 1990).

The transfer of the TINA vision system model matcher [28,29] to the fast parallel vision system MARVIN [10] exploited ideas central to the YASA project. The combinatorial complexity of the search problem was much reduced by restructuring the object model into its characteristic views each with its own particular set of focus features. Furthermore, since the MARVIN system is able to conduct the simplified model matching task for each characteristic view in parallel, a considerable speed up is obtained.

One spin-off of the YASA project was a psychophysical study of human object recognition conducted in AIVRU by Langdon (a postgraduate student of Mayhew and Frisby). A report of that work, which as it developed became as much a pursuit of mechanisms of mental rotation as it did of canonical view-based object representations, is included here [27] as a fitting tombstone for the YASA project.

The work in the 3D Model-based Vision Project culminated in two systems integrating research within but not between sites. This reflects the nature of the collaboration enjoyed by the consortium: a loose club of communicating but autonomous modules (the platoon or Vietcong model; contrast with the nexus of interdependencies or pack of cards model).

Sheffield chose to demonstrate the results of the research in the three sub-projects (PMF, 2.5D and 3D Model-based Vision) by using stereo vision to solve the 'pick and place task' cliche [24, 25]. The system of component modules was called TINA for reasons which can no longer be remembered but *T*hat *I*s *N*o *A*nswer and *T*his *I*s *N*o *A*cronym seem to capture some of the flavour of the thinking at the time. Given the present predilection of so many industrial automation engineers to design vision out of their production lines, we might with hindsight suggest that *T*here *I*s *N*o *A*pplication might be a more appropriate interpretation. However, if flexibility is to be the touchstone of the future factory, then the best interpretation might be *T*here *I*s *N*o *A*lternative.

TINA has now been completely rewritten (almost entirely by Pollard) and provides the basis of TINATOOL, a very extensive integrated vision research environment. TINATOOL has been ported to several research sites. This experience taught us first-hand the oft-repeated warning that porting large bodies of software can be an extremely time-consuming and frustrating task as the attendant responsibilities become manifest. It should not be undertaken lightly, particularly in a consortium devoted primarily to basic research.

B (contd) Notes Added by Fisher

The work at Edinburgh started with two tracks. The first track investigated methods for applying model-based vision methods in situations where most object identities and positions are known, such as in a typical robot workcell, containing the known robot, gantry, feeders, jigs and workpiece. The remaining objects to be visually analysed may be a dropped part, or a part with an unknown orientation. A CSG model of a known scene is used by

ROBMOD (Cameron, 1984) to deduce a wire-frame model of the visible 3D edges, in an off-line process. To do this, we extended the ROBMOD body modeler to deduce boundary representations from the Constructive Solid Geometry object. ROBMOD was also extended to produce an annotated visible edge description, by analysing the object visibility from a given viewer position. This produced a list of the visible portions of the 3D object edges, including extremal boundaries.

The wire-frame edges are then matched to 3D data edges, such as those obtained from a stereo camera system. We used a set of simple position constraints to verify the matches close location, close orientation, data edge within predicted model edge, etc.). Because the 3D positions of the model and data edges should be identical, fast matching is possible, and we achieved a 1 second verification time [25].

The other track investigated model-based object recognition and location, based on surface patch evidence, such as would be represented in the REV graph. During the early part of the project, work was spent investigating the IMAGINE I system (developed for Fisher's PhD thesis, now reported as a monograph - Fisher, 1989a). This work pointed out particular problems in the areas of model representation and geometric reasoning.

As a result of the evaluations, the SMS object representation scheme (Fisher, 1987a) was designed and implemented [24]. This was a surface-based modeler to allow connecting curved surfaces, surfaces with holes, degrees of freedom and a greater variety of surface types. As any model feature could be described using expressions involving variables, models could have deformable parametric shapes (but not variable structure). A generalisation hierarchy was also included, to allow scale dependent representations. Because the volumetric features did not represent well the significant features of the object, such as might be used to suggest or confirm identity, second-order volumetric primitives were added to the models (Fisher, 1987b,c).

Since a considerable portion of model matching time was spent in appearance prediction (to derive feature visibility and self-occlusion relationships) the SMS models were designed to include a visibility submodel, listing the features visible from salient viewpoints and new viewpoint dependent features, such as extremal boundaries. This idea linked closely with the proposed YASA representation.

Another observation from the IMAGINE I system was that the geometric reasoning system was insufficiently powerful for complex problems. By examining previous 3D vision systems, Orr and Fisher (1987) identified the key geometric reasoning functions, needed for vision applications, and used these to guide the development of a new interval arithmetic geometric reasoner based on propagating bounds on quantities through a parallel network (Fisher, 1988; [23]). The network approach allowed handling of data errors, model variations including degrees of freedom, incremental position constraints, and a priori scene constraints. We observed that the forms of the algebraic position constraints tended to be few and repeated often, and hence standard subnetwork modules could be developed, with instances allocated when new position constraints were identified.

Though research on the use of these networks is continuing (e.g. Fisher, 1989b; Fisher and Orr, in press) problems overcome by the new technique included the weak bounding of transformed parameter estimates and partially constrained variables, and the representation and use of constraints not aligned with the parameter coordinate axes. The network also has the potential for large scale parallel evaluation. This is important because about one-third of the processing time in the scene analyses was spent doing geometric reasoning.

The IMAGINE I system showed the importance of having a model invocation process (Fisher, 1989a) to identify quickly candidate models from the model database. The work undertaken as this part of the project resulted in: (1) extensions for including inhibiting relationships, which produced considerably improved invocation behaviour, (2) extensions for using binary feature evidence (i.e. spatial relationships between two features, such as relative distance, orientation or size), and (3) implementing the model invocation process in an explicit value-passing network. This work is still continuing, and is awaiting more complete experimentation before having a proper report.

In addition, Paechter (1987) observed that many database objects required similar properties, such as requiring that principal curvatures must be zero. From this, he introduced a hierarchy of low-level symbolic description types, such as 'planar', for the example above. This did not change the network competence, but dramatically simplified the definition of object properties. Another major class of improvements were for a family of evidence evaluation functions, for example when a property's exact value is unimportant provided it is positive. The evidence evaluation functions were changed to have gaussian form, which linked the properties more closely to their statistical characterization. A more uniform evidence integration function based on an harmonic mean was introduced, which then allowed subcomponent evidence to be uniformly integrated with property evidence.

To match models to surface patches, one has to overcome occlusion and fragmentation, organise these features into groups, describe properties of the objects, select models, pair model features to the data features, estimate object positions and reason about missing features. To do these, the IMAGINE II system [26] was designed. Based on this, a framework was built for undertaking these actions, and developed sufficiently to demonstrate one complete recognition before the end of the grant period. The data surfaces were correctly paired to the model surfaces, and the substructure hierarchy developed correctly. Object position was estimated accurately enough to be barely distinguishable from the perfectly correct position. The successfully recognized object was an oil bottle using laser range data[3]. This recognition, by itself, was not significant, but was a promising first step, in that it was of a non-polyhedral object.

REFERENCES

Attneave, F. (1954) Some Information aspects of visual perception. *Psychological Review* **61** 183-193.

[3] Editors' Note. The *Needles* stereo algorithm (see [20]) would now be able to provide similar dense range data from visual images but it was not available at the ime Fisher's work required it.

Aylett, J. C., Fisher, R. B., and Fothergill, A. P., (1988) Predictive computer vision for robotic assembly. *Journal of Intelligent and Robotic Systems* **1** 185-201.

Baumgart, B.G. (1972) Winged edge polyhedral representation, STAN-CS-320, AIM-179, Stanford AI Lab.

Binford, T.O. (1982) Survey of model-based image analysis systems. *International Journal of Robotics Research* **1** 18-64.

Bolles, R.C., Howard, P. and Hannah, M.J. (1983) Three dimensional part orientation system. Proc. *International Joint Conference on AI*, 1116-1120.

Brooks, R.A. (1981) Symbolic reasoning among 3-D models and 2-D images. *Artificial Intelligence* **17** 285-348.

Brown, C. M. (1982) PADL-2: A technical summary. *IEEE Computer Graphics* **2** 69-85.

Cameron, S. A. (1984) *Model solids in motion.* PhD Thesis, Department of Artificial Intelligence, University of Edinburgh.

Fisher, R. B. (1987a) SMS: A suggestive modeling system for object recognition. *Image and Vision Computing* **5** 98-104.

Fisher, R. B. (1987b) Model invocation for three dimensional scene understanding. *Proc. Int. Joint Conf. AI*, 805-807.

Fisher, R. B. (1987c) Modeling second-order volumetric features. *Proc. 3rd Alvey Vision Conference*, 79-86, Cambridge.

Fisher, R. B. (1988) Solving geometric constraints in a parallel network. *Image and Vision Computing* **6** 1988.

Fisher, R. B. (1989a) *From surfaces to objects: computer vision and three dimensional scene analysis.* Chichester: John Wiley & Sons.

Fisher, R. B. (1989b) Experiments with a network-based geometric reasoning engine. Proc. *International Joint Conference on AI*, 1632-1628.

Fisher, R. B. and Orr, M. J. (in press) Geometric reasoning in a parallel network. *International Journal of Robotics Research.*

Hrechanyck, L.M. and Ballard, D.H. (1982) A connectionist model of form perception. *IEEE*, 44-52.

Koenderink, J.J. and Van Doorn, A.J. (1979) The internal representation of solid shape with respect to vision. *Biological Cybernetics* **32** 211-216.

Marr, D. and H.K. Nishihara (1978) Representation and recognition of the spatial organisation of three dimensional shapes. *Proc. of the Royal Society of London, B.* **200**, 269-294.

Orr, M. J. L. and Fisher, R. B. (1987) Geometric reasoning for computer vision. *Image and Vision Computing* **5** 233-238.

Paechter, B. (1987) *A new look at model invocation with special regard to supertype hierarchies,* MSc Dissertation, Department of Artificial Intelligence, University of Edinburgh.

Perrett, D.I. and Harries, M.H. (1988) Characteristic views and the visual inspection of simple faceted and smooth objects: tetrahedra and potatoes. *Perception* **17** 703-720.

Rieger, J.H. (1987) On the classification of views of piecewise smooth objects. *Image and Vision Computing* **5** 91-97.

Rieger, J.H. (1990) The geometry of view space of opaque objects bounded by smooth surfaces. *Artificial Intelligence* (in press).

Requicha, A.A.A.G. (1980) Representations for rigid solids: theory, methods and systems. *Computing Surveys* **12** 437-464.

Requicha, A.A.A.G. and Voelcker, H.B. (1982) Solid modelling: a historical survey. *IEEE Computer Graphics* **2** 9-26.