III 2.5D SKETCH PROJECT

Introduction by the Editors

A THE GRANT PROPOSAL (Written 1983)

1 OBJECTIVES

The objective of the 2.5D Sketch Project is to develop methods for deriving from the depth map delivered by the PMF Project a representation of the 3D structure of the visible surfaces in the scene. The 2.5D Sketch is conceived here as an intermediate-level representation serving as the input representation for the 3D Model-Based Vision Project as well as being able to support in its own right various lower-level tasks such as some forms of trajectory planning, robot guidance, grasp planning etc.

It is proposed that the output representation of the 2.5D Sketch Project be a relational structure describing the vertices, edges, and faces in the scene and their adjacency and connectivity relationships. This structure is in some ways analogous to the winged edge shape representation used by Baumgart (1972) to describe polyhedral objects. That representation is basically a pointer-record structure in which the nodes/records correspond to the elements (vertices, edges, regions) and the arcs/pointers express the adjacency and connectivity relationships. Such a relational structure is extendable, and quite general; it is envisaged that the records will contain fields containing not only the geometrical coordinates of the entities, but also, as far as can be recovered, both their qualitative and quantitative 3D descriptions. It would be premature to attempt to specify further the details of the representation to be used for the 2.5D Sketch Project, and wherever the term 'winged edge' is used hereafter it is to be understood as a generic label rather than a commitment to a precise form of relational structure.

2 LOCATION AND DESCRIPTION OF SURFACE DEPTH AND SURFACE ORIENTATION DISCONTINUITIES

The basis of the proposed 2.5D scene description is the discovery and identification of the surface regions bounded by depth and surface orientation discontinuities. The labelling and classification of surface orientation and depth discontinuities is a necessary precursor to grouping the scene into regions. An important stage in the labelling and description of the regions in the scene is to identify and label the ground plane. If the stereo camera geometry is known (as it will be) the labelling of the ground contact edges of the objects in the scene will in general be relatively straightforward.

The information delivered by PMF is local, sparse and noisy. Its range data output is tied to image description entities such as zero crossing segments, and although in many cases these derive from the projection of surface orientation discontinuities, boundaries and occlusions, the sign of the edge, and the direction of the occluding-occluded relationship is not explicitly represented by PMF. There are several strategies for the recovery of this information which the 2.5D Sketch Project will evaluate.

One possibility is to use a stage of surface interpolation either directly integrated with the stereo correspondence process or else operating over its output. Ideally the interpolation process should be self-monitoring and break at surface depth and orientation discontinuities, otherwise a subsequent 'edge detection' process is required to detect and label the surface depth and orientation discontinuities. It is anticipated that surface interpolation will be a substantial part of the 2.5D Sketch Project with detailed proposals described below.

Alternatively, because the profile of the local stereo disambiguating support (particularly if augmented by some form of grey level correlation working within the constraints of edge based stereo; see the PMF Project) will often constrain the choice of the sidedness of occlusions, it should be possible to obviate the necessity of an independent stage of surface interpolation in many cases. And in others, this information could be used to help direct the interpolation.

Another possibility is to exploit the propagation of 3D information from cues in the 2D image or primal sketch features as in classical line labelling schemes (Kanade, 1981; Draper, 1981; for a review, Binford, 1982) though the resolution of the low level visual processing needed for such a scheme may be prohibitive.

A robust system will probably need the intelligent conjunction of all these and possibly other strategies to recover surface shape descriptions. It will also probably need to incorporate information about the illumination characteristics of the scene, eg for the principled treatment of specularities and shadows.

3 QUALITATIVE META-FEATURE DESCRIPTIONS

It is the intention of the proposal to develop procedures to recover 3D features and their relationships from the depth map of the scene using only image/depth map descriptive processes and without recourse to higher level semantics. The obvious basis for the relationship between features is their connectivity, eg surface regions connected across edges derived from surface orientation discontinuities, vertices connected along such edges etc. It is likely however that other 3D topological relationships will also prove useful descriptors. For example, if an object contains a row of holes within a face then the colinear grouping over the individual image features provides a gestalt that can be exploited in the model invocation and recognition process. We thus propose that if a global feature can be reliably and cheaply computed from the depth map, and if it is a potentially useful index into the 3D model catalogue capable of guiding the initial stages of the recognition process, then such gestalts should be part of the 2.5D Sketch description. Typically such meta-features would be colinearities and parallel lines but symmetries might also be included as might simple prototypical shapes. These meta-features could also include useful global metrical information such as an estimate of an object's size.

4 SEGMENTATION AND DESCRIPTION OF COMPLEX SMOOTH SURFACE REGIONS

If a complete depth map of a surface region has been recovered from the stereo range data, then it may be desirable to recover a description of the surface shape within the region boundaries. In the proposed representation a surface region is defined as a region of the depth map enclosed by surface orientation or depth discontinuities, or occlusion or extremal boundaries. A surface region is thus C1 continuous and may be either simple or complex. We describe as simple a surface region in which the magnitudes of the principal curvatures may vary over the surface but in which the signs of the principal curvatures do not. Thus a simple surface contains no inflexions of curvature and in these terms the quadratic surfaces are simple. It is proposed that complex surfaces can be segmented into regions along the lines of curvature corresponding to the extreme and/or the inflexions of the principal curvatures (Brady, 1982; Hoffman, 1983). If the result of 'interpolation' over the segmentation is itself a complex surface, then a further stage of segmentation can be applied and so on until a simple surface is obtained. This successive segmentation and 'interpolation' over an increasingly coarser (but not spatially blurred) sample of the surface data points can be used to recover a description of the surface over its natural scales. The quotes around interpolation are to suggest that the qualitative description of the surface recovered from the straight line polygonal hull of the segmentation control points may be sufficient and so avoid the cost of repeated full interpolations.

5 SURFACE INTERPOLATION

Reference was made above to the desirability of developing methods for interpolating surfaces between the sparse depth map delivered by PMF. This topic is explored here in more depth and the work that Blake plans in this area is described.

The interpolation process will be considered in close conjunction with the objective of labelling depth discontinuities and other surface properties. Also, the interpolated surface must be consistent with the original intensity data, in the sense that where the intensity distribution can be wholly or partly predicted from 3D surface structure (and in particular from surface discontinuities) such predictions should agree with the original intensity data.

5.1 Surface Interpolation and Labelling Discontinuities

It is well known that interpolation of sparsely distributed values in an array, to form a smooth 2D function, can be achieved in a highly parallel manner. For example, stretching an elastic membrane over a fixed wire frame is just such an interpolation and if the form z(x,y) is made at discrete points over a regular x,y grid, then at each point the height value is successively replaced by the average of the heights of its near neighbours. On the wire frame itself, height values remain fixed. This replacement continues until height values at all points have converged. The algorithm is local in that the average computed at each point involves only the point's near neighbours and parallel because averages can be computed simultaneously at all grid points. Given a suitable parallel machine, such an algorithm can be executed rapidly; this is important in a vision system for which real-time performance is required. This algorithm is a simple example of what are generally called 'relaxation algorithms'. The simple interpolation algorithm just described can be regarded as performing an optimisation over an x,y grid to minimise the energy (elastic potential energy) of the membrane. A different choice of energy function results in a different sort of interpolation and in many cases it is still possible to minimise the energy by relaxation (Ullman, 1979). For instance Terzopoulos (1982) uses a more complex energy function - for interpolating stereo disparity data - that represents the energy of a thin plate rather than a membrane. Such a plate resists flexion and torsion and therefore its surface remains free of fractures and creases (it is continuously differentiable). In general the visible surface is not continuous and the interpolation procedure has no ability to locate discontinuities in the surface and its gradient. Rather it requires the location of discontinuities to be determined in advance so that the interpolation process can be inhibited across them. Thus interpolation proceeds only within the boundaries defined by the discontinuities, resulting in a piecewise continuous surface.

However, discontinuities are a very important part of the surface description and it is important to locate them reliably. It is attractive to try and incorporate discontinuity detection into the interpolation process. Intuitively this is plausible if one thinks of trying to fit a smooth surface everywhere, but allowing discontinuities to form where the 'strain' on the surface becomes too great. Blake (1983a, 1983b) describes how this can be done for the somewhat restricted case of discontinuities in a piecewise constant surface map. It is achieved by imposing 'weak continuity constraints' (Hinton, 1977, introduced weak constraints in a rather different context) to express the expectation that the surface 'varies smoothly almost everywhere' (Marr, 1982). Such constraints are only broken where the data to which the surface is being fitted (eg raw disparity data) forces a discontinuity. The ease with which constraints may be broken is controlled by a constant (the 'penalty constant'). A high penalty results in few discontinuities and a low one produces many - giving greater detail - so that the scale of the segmentation of the surface into continuous pieces is controlled by the penalty constant. The scheme described in Blake (1983a,b) has no ability to detect gradient discontinuities, or to cope with extensive smooth surfaces at large angles of tilt or large curvature, and is therefore unable to construct visible surface maps for general scenes. However it works well within its limitations - it has been tested finding discontinuities in image intensity data (which is not sparse) - and promises to be susceptible to the necessary generalisation by incorporating an energy function like that of Terzopoulos and modifying the algorithm to work with sparse data.

5.2 Consistency of the Surface Map with Intensity Data

Information from left and right images is combined in stereopsis to produce the disparity map, but there may be still further information in the intensity images that is relevant to the construction of the surface map. It is clear that the distribution of intensity constrains the surface map (via the image irradiance equation) and Ikeuchi and Horn (1981) show how the shape of a continuous surface patch can be recovered (by a relaxation algorithm) from intensity data. In practice, of course, this is likely to be complicated by lack of precise knowledge of the ambient illumination. Intensity data also contain powerful clues about the location and type of surface and illumination discontinuities (Barrow and Tenenbaum, 1978; Witkin, 1982). Stereopsis can make partial use of such information, eg in the figural continuity constraint (Mayhew and Frisby, 1981). Further use of the information could be made by specifically enabling discontinuity formation, during construction of the surface map, where there are discontinuities in intensity of the appropriate type. For example, this would apply where the tangency condition (Barrow and Tenenbaum, 1978) is satisfied, indicating the presence of an occluding edge, but not along surface markings.

5.3 Summary of Work Proposed on Surface Interpolation

It is proposed, starting from the Pollard, Mayhew and Frisby stereo algorithm (PMF), to design and implement an algorithm for surface reconstruction. This will draw on work already done on interpolation by relaxation (Terzopoulos, 1982) and on labelling discontinuities by the use of weak constraints (Blake, 1983b) combined and extended to maintain consistency of the surface map with intensity data. The work will include investigating the potential for fast multilevel processing (processing at a variety of coarse and fine resolutions, as in Terzopoulos, 1982, and Glaser, 1983) and the applicability of parallel, statistical methods for solving optimisation problems (Metropolis et al, 1953; Hinton and Sejnowski, 1983). As the work develops consideration will be given to suitable parallel architectures for efficient implementation of the surface reconstruction algorithms.

B WHAT REALLY HAPPENED?

The 2.5D Sketch Project was soon redefined and renamed as the REV Graph Project (***R***egions, ***E***dges, ***V***ertices), and enlarged to include explicitly the Wire Frame Completion project (WFC), whose goal was to recover the 3D wireframe scene description from edge-based stereo.

The REV Graph proposal (Blake and Mayhew, 1987) was a design for a low level image processing architecture owing much to the blackboard metaphor (Hayes-Roth, 1985), the notion of intrinsic images (Barrow and Tenenbaum, 1978), and visual routines (Ullman, 1982). Some of the component modules or 'knowledge sources' were eventually developed and are described in the papers that follow. One particular KS was to identify specularities. Specularities may on the one hand be regarded as a source of noise, but if correctly identified may also provide a source of information concerning the curvature of the surface. Brelstaff, a graduate student of Blake worked on the problems both of identifying [17] and exploiting [18] specular reflections. The result was an augmented version of Sheffield's PMF that excised specular features to avoid matching errors arising from their violation of epipolar constraints.

Stereoscopic disparities of specularities were actually used to determine the curvature of the underlying surface by Zisserman , Giblin and Blake (1989). It has recently been shown that curvature perception in human stereoscopic vision is similarly influenced by specularities (Blake and Bulthoff 1990).

Another knowledge source specified in Blake and Mayhew (1987) was a stereo module specialised for dealing with highly textured surfaces. This culminated in the PhD work of McLauchlan in AIVRU on the Needles stereo algorithm which is described in paper [20].

The REV Graph architecture was explored only in prototype form in the context of the WFC Project. The ANIT system (Mayhew, 1989; Booth and Mayhew, 1989), which has explored some of the good ideas from Brooks' (1986) subsumption architecture, has its antecedents in this work, and is currently being transferred to a transputer environment.

The WFC Project was originally implemented in Lisp by Bowen and metamorphosed into the Consistency Maintenance System (CMS) reported in [19]. This built on the ideas of Herman and Kanade (1986) for wire frame model acquisition and those of De Kleer (1986) for consistency maintenance. The system was never fully integrated into AIVRU's TINA vision processing environment for tedious technical reasons related to the evolution of increasing incompatibility between Sun graphics, Lisp, C and Unix with each generation of the operating system as the project continued.

The WFC Project reached its zenith in the development of Geomstat by Porrill [12], a geometrical reasoning module. Exploiting Gauss-Markov optimal estimation techniques, Geomstat is a system for the integration of edge and curve descriptions (and their associated error models) to recover 2D and 3D geometric descriptions of surfaces and their intersections. Recent attempts to integrate Geomstat with a usable Lisp-based intelligent front-end have again struggled against the continuing inadequacy of the Lisp-C-graphics interface on Sun workstations.

The work on qualitative meta-feature descriptions was begun with Ian Graydon of GEC in a project using the Hough transform to find rows of blobs. The project then metamorphosed into a method for recognising planar objects using methods derived from the work of Bolles and Fischler (1981) which used focus features to devise a strategy to minimise the computational overhead of the maximal clique algorithm. This work was reported to the 1986 Alvey Vision Conference (for which no formal proceedings were published - hence no report is included here; Graydon left the project before a formal paper could be prepared for journal publication).

B (contd) Notes Provided by Blake

The remaining major component of the 2.5D sketch project concerned the recovery of surface discontinuities and descriptions. At the beginning of the project, Blake showed that the continuous surface reconstruction scheme of Grimson (1982), elaborated by Terzopoulos (1983), was flawed. It was unsuitable for what is now sometimes called *Active Vision* because it lacked viewpoint invariance [13] . The resulting instability of reconstructed surfaces was most pronounced when stereoscopic features were sparse, precisely the conditions under which reconstruction was supposed to be most useful! At the same time the emphasis of the problem shifted. What was the purpose of extrapolating surfaces through large tracts of empty space? Instead, the primary role of surface reconstruction in vision seemed to be the recovery of surface discontinuities on textured surfaces, where they would not have been visible monocularly.

Blake and Zisserman decided to concentrate on the reconstruction of surfaces complete with their discontinuities.

In the absence of established methods in the spline literature, it was natural to combine the idea of 'weak continuity' constraints with viewpoint invariance [14]. The earlier *Graduated Non-Convexity* (GNC) algorithm (Blake 1983a,b) was extended to deal with the 'weak membrane'. It turned out to be very effective experimentally, efficient compared with stochastic techniques [15], and provably correct for a significant class of signals. This led to the investigation of a whole family of reconstruction problems involving discontinuities, subsequently published in a book (Blake and Zisserman 1987). Variational analysis established some remarkable properties, in particular the stability of the weak membrane in noise and in scale-space. Unlike the Gaussian case (Witkin 1983), the new scale-space proved to be 'uniform' so that discontinuities at different scales were in perfect registration [16]. Around the same time, Mumford and Shah (1985) made major contributions to the variational analysis which were combined with Blake and Zisserman's own results in the book. At that time, analysis was completed only for the case of one-dimensional signals and hence, of course, for two dimensional signals with translational symmetry. Mathematicians are, even now, struggling with the full two-dimensional problem.

REFERENCES

Barrow, H.G. & Tenenbaum, J.M. (1978) Recovering intrinsic scene characteristics from images. In *Computer Vision Systems*, A.R. Hanson & Riseman, E.M. (eds.), Academic Press: New York. Pp. 3-26.

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 understanding systems. *Internat. J. Robotics Res.* **1** 18-62.

Blake, A. (1983a) The least-disturbance principle and weak constraints. *Pattern Recognition Letters* **1** 393-399.

Blake, A. (1983b) *Parallel computation in low-level vision.* PhD Thesis, University of Edinburgh.

Blake, A., and Mayhew, J. E. W. (1987) Alvey 2.5D sketch project: Proposed structure for a development system. *AI Vision Research Unit Memo No 9*, University of Sheffield.

Blake, A. and Zisserman, A. (1987) *Visual reconstruction.* MIT Press: Cambridge, Mass.

Blake, A. and Bulthoff, H. H. (1990) Does the brain know the physics of specular reflection? *Nature* **343** 165-168.

Bolles R. C. and Fischler M. A. (1981) A RANSAC-based approach to model fitting and its applications to finding cylinders in range data. *Proc. Int. Joint Conf. AI 1981.* 637-643. Vancouver, B. C., Canada.

Booth, C. and Mayhew, J.E.W. (1989). Implementing a behavioural decomposition. Poster at *Intelligent Autonomous Systems Two*, Amsterdam, December 1989.

Brady, M. (1982) Criteria for representations of shape. *MIT AI Laboratory Memo.*

Brooks, R. A. (1986) A robust layered control system for a mobile robot. IEEE *Journal of Robotics and Automation* **2** 14-23.

De Kleer, J. (1986) An assumption based TMS. *Artificial Intelligence* **28** 127.

Draper, S. W. (1981) *Reasoning about depth in line-drawing interpretation.* PhD Thesis, University of Sussex.

Glaser, F. (1983) Multilevel relaxation in low-level computer vision. In *Multiresolution image processing and analysis*, A. Rosenfeld (ed.), Springer-Verlag.

Grimson, W. E. L. (1982) *From images to surfaces.* MIT Press: Cambridge, USA.

Hayes-Roth, B. (1985) A blackboard architecture for control. *Artificial Intelligence* **26** 251.

Herman, M. and Kanade, T. (1986) Incremental reconstruction of 3D scenes from multiple, complex images. *Artificial Intelligence* **30** 289.

Hinton, G. H. (1977) *Relaxation and its role in vision.* PhD Thesis, University of Edinburgh.

Hinton, G. E. and Sejnowski, T. J. (1983) Analysing cooperative computation. *Proc. 5th Conf. Cognitive science*, Rochester, New York.

Hoffman, D.D. (1983) *Representing shape for visual recognition.* PhD Thesis, Dept. of Psychology, MIT, Cambridge, Mass.

Ikeuchi, K. and Horn, B. K. P. (1981) Numerical shape from shading and occluding boundaries. *Artificial Intelligence* **17** 141-184.

Kanade, T. (1981) Recovery of the three dimensional shape of an object from a single view. *Artificial Intelligence* **17** 409-460.

Marr, D. (1982) *Vision.* San Francisco: W H Freeman & Co.

Mayhew, J.E.W. & J.P. Frisby (1981) Towards a computational and psychophysical theory of stereopsis. *Artificial Intelligence* **17** 349-385.

Mayhew, J. E. W. (1989) The ANIT system: Architecture for *N*avigation and *I*ntelligent *T*racking. *AI Vision Research Unit Memo Memo No 48*, University of Sheffield.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953) Equation of state calculations by fast computing machines. *Journal of Chemical Physics* **6** 1087.

Mumford, D. and Shah, J. (1985) Boundary detection by minimising functionals. *Proc. IEEE CVPR Conf.* **22.**

Terzopoulos, D. (1982) Multilevel reconstruction of visual surfaces. *MIT AI Lab Memo 671*, Cambridge, Mass.

Terzopoulos, D. (1983) The role of constraints and discontinuities in visible-surface reconstruction. *Proc Int. Joint Conf. AI 1983* 1073-1077.

Ullman, S. (1979) Relaxation and constrained optimization by local processes. *Computer Graphics and Image Processing* **10** 115-125.

Ullman, S. (1984) Visual routines. *Cognition.* **18** 97-160.

Witkin, A.P. (1982) Intensity based edge classification. *Proc. AAAI Conference* 36-41.

Witkin, A. P. (1983) Scale-space filtering. *Proc. Int. Joint Conf. AI 1983* 1019-1022.

Zisserman, A., Giblin, P. and Blake, A. (1989) The available information to a moving observer from specularities. *Image and Vision Computing* **7** 38-42.