[20] Location and Description of Textured Surfaces Using Stereo Vision

Philip F McLauchlan, John E W Mayhew and John P Frisby

AI Vision Research Unit
University of Sheffield, Sheffield S10 2TN, UK

We describe a stereo algorithm, called Needles, specialised to deal with smooth textured surfaces. Constraints of local surface smoothness and global surface continuity are used to solve the correspondence problem. The algorithm is edge based. First the left image is divided into square patches, and a disparity histogram of the potential edge matches is constructed in each patch. Above-threshold peaks in the histogram are passed into a Hough transform, which fits a plane to a subset of the potential matches lying around the peak, forming local hypotheses for the range and orientation of the visual surface along with the edge matches. Next, hypotheses in adjacent, overlapping patches are connected if they share enough common matches. A region growing procedure locates large areas of mutually connected hypotheses, corresponding to continuous, possibly overlapping surfaces. When surfaces overlap, the largest one is chosen. Needles has been implemented on a Sun workstation and a Transputer network. Results are presented for two stereopairs, and compared with physical measurements.

In approaches to solving the stereo correspondence problem in computer vision to date, little explicit attention has been paid to the special problems posed by highly textured surfaces. Edge maps associated with such surfaces tend to be dense, fragmentary and noisy. These characteristics make the correspondence problem much harder to solve than normal. This paper demonstrates an algorithm, called Needles, which can overcome these difficulties for smooth surfaces by implementing a strong form of the smoothness constraint widely used in stereo algorithms.

One of the advantages of Needles is that it generates a visual surface description directly as part of the matching process. In many feature-based stereo algorithms, some species of smoothness is exploited in the form of mutual support propagated between matches that could lie on a "smooth" surface. In subsequent surface reconstruction, as proposed by Grimson [5] and Terzopoulos [15] for example, the information used in the application of the smoothness constraint at the matching stage is then discarded: a thin plate surface is fitted to all matches whether or not they supported each other. In contrast, although Needles discards most of the potential matches by application of a local smoothness constraint, final matching decisions are postponed until the stage at which a surface description is selected. Unlike algorithms that solve the correspondence problem by propagation of local constraints, such as Barnard and Thomson's [1] and PMF [10], Needles uses a combination of local and global constraints, the latter being based on continuous whole surfaces rather than local patches.

The general approach of integrating matching and surface reconstruction has been used previously, for example by Boult and Chen [2] and by Hoff and Ahuja [6]. However, Needles differs in using a global region-growing procedure to link neighbouring local patches of potential smooth-surface matches if they share a sufficient number of matches in their region of overlap. Final disambiguation is applied to the continuous surfaces formed in this way. Like [2] and [6], Needles integrates surface reconstruction and surface discontinuity detection. The global disambiguation mechanism distinguishes Needles from other algorithms that use a locally planar model of disparity, such as [6] and that of Otto and Chau [9].

1   The Needles Algorithm

The Needles algorithm is feature based, using at present edgels produced by the Canny edge detector [4]. It is assumed that at the scale at which the algorithm is applied (defined by the image patch size, see below) the variation of the visual surface from a plane is small relative to its extent. This assumption provides a very strong constraint on the possible edge matches. A brief summary of the algorithm is as follows: one image (the left) is divided into small square overlapping patches. In each patch a histogram of the disparities of the potential matches is constructed. Peaks in the histogram provide hypotheses for the disparity of the visual surface in the patch. A Hough transform then selects from the potential matches near each hypothetical disparity a set of them all of which lie near a plane. The other potential matches are rejected. Sets of matches from adjacent image patches that contain enough matches in common are labelled as connected (i.e. as part of the same surface). A region growing procedure finds large regions of mutually connected sets of matches. Where regions overlap the strongest region (in the sense defined in section 1.4 below) wins, and its matches are selected.

Needles thus uses local (within patch) matching constraints to form hypotheses for the visual surface in each patch. The incorrect hypotheses are eliminated using *global* surface connectivity information, i.e. each surface is located as a whole. Surface smoothness is used in two ways: a local smoothness constraint to generate local surface hypotheses, and a surface continuity constraint to make explicit the connectivity of the local hypotheses.
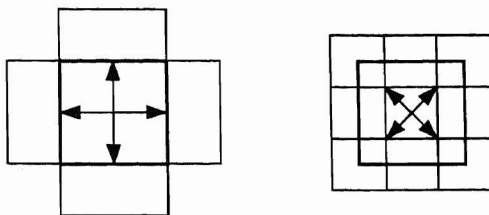
Figure 1: The four lateral (left) and four diagonal (right) neighbours of an image patch. The central patch is shown in bold. The arrows mark the differences in position of the adjacent patches with respect to the central patch.

1.1 Preprocessing

The left image of the stereopair is divided into overlapping square patches of width $H = 32$ pixels. The patches are arranged in a grid so that diagonally adjacent patches overlap by 3/4 of their length in each direction, while laterally adjacent patches overlap in half their area. Each patch thus has eight neighbours as shown in figure 1. A simple test on texture density rejects an image patch if the number of edges within it is less than a threshold[1] $2H$.

The edge positions are *rectified* to the positions they would have been in had the camera image planes been parallel[2]. This is done using camera parameters obtained from Tsai's camera calibration method [16]. Edge detection and rectification take place within AIVRU's TINATOOL stereo vision environment [12]. Corresponding edges in the two images are now assumed to lie in the same image raster. Given a pair of edges with rectified positions (x_l, y) and (x_r, y), disparity is defined as $d = x_r - x_l$.

Each edge pair lying in the same raster must satisfy five *compatibility* conditions in order to be accepted as a potential match. The conditions are:

1. The disparity of the edge pair must lie within a (large) initial range extending $0.375s$ on either side of the convergent point of the optic axes of the cameras, where s is the size of the image in pixels.

2. The contrasts of the edges are compared. If the ratio of the larger to the smaller is greater than a threshold, set at 4, the pair is excluded.

3. Neither edge can have an orientation within $5°$ of horizontal. Near horizontal edges give rise to large disparity measurement errors.

4. The orientation of the edges must be the same side of horizontal, i.e. an edge marking a boundary between a light region on the left and a dark region on the right can only match with another edge of the same type. This is an analogue of the contrast sign rule characterising human vision (but see [11]).

---

[1] Note: due to lack of space, full explanations are not given for the values of all the parameters quoted, but they are given in [8]. The quoted values have been found to give good results on all the images so far tested.

[2] This corresponds to a rotation of the cameras about their optical centres to bring the image planes into alignment.

5. Edge orientations correspond to the orientations of boundaries in the images, which may be due to object boundaries, surface texture etc. For a pair of edges to be matched it must be feasible for the orientations to be projections of a boundary in the world. Since Needles imposes a disparity gradient limit on the surfaces it finds (as explained below) it is reasonable to impose a limit on the disparity gradient of the line in disparity space (x, y, d) formed by back-projecting the edge orientations. This is set to 1.

1.2 Disparity Histogramming

A local disparity histogram [14] is constructed for each square patch. The disparity range is divided into blocks of size $0.2H$, and an accumulator assigned to each block. Each compatible edge pair contributes a vote to the corresponding disparity block. The magnitude of the vote is the *weight* assigned to the left edge of the pair. This is an integer dependent on the position of the edge within the square patch. The weighting function is a pyramid with its peak at the centre of the patch. The weights are used throughout the algorithm, and the increased weight assigned to central edges means than matches are less inclined to congregate on one side of a patch. This gives rise to more reliable connections between adjacent patches.

The histogram is smoothed by Gaussian convolution, using a mask with $\sigma = 1.5$ blocks. The peaks in the smoothed histogram are then thresholded. The threshold is set to $0.1W$, where W is the sum of the weights of the left edges in the square patch. Each peak above the threshold is localised by fitting a quadratic to the histogram accumulator values at and on either side of the peak value. The maximum of the quadratic is taken to be the disparity at the peak, which is then passed into the next stage, plane fitting.

1.3 Planar Patch Fitting by Hough Transform Method

The transformation between disparity space and world space preserves planes. The plane fitting can hence be done in disparity space. For each peak in the disparity histogram an attempt is made to fit a plane through the disparity points lying near the peak. A large number of these points will be incorrect, so direct fitting, by least squares for example, would not work. A Hough transform is used to select a large subset of the points which lie near a plane. This point set, along with the plane parameters, constitutes a local *surface hypothesis*.

For each image patch the origin of the left image coordinate system (x, y) is reset to the centre of the square. The equation of a plane in disparity space can be written as

$$d = ax + by + c \qquad (1)$$

where a, b and c are constant (a and b are the disparity gradients in the x and y directions respectively). For a given point (x, y, d), eq. 1 describes a plane in parameter space (a, b, c) defining the set of values of a, b and c that give rise to a plane in disparity space passing through the

point (x, y, d). Points lying on the same plane in disparity space define planes in parameter space which meet at a single point. The problem is to find that point. The standard Hough transform approach is to divide parameter space into blocks in each direction. For each point (x, y, d), and each block in parameter space that the plane in eq. 1 passes through, an accumulator assigned to the block is incremented. At the end the block whose accumulator that received the most votes is the best planar fit to the data.

The Fast Hough Transform (FHT)

The above method has two main drawbacks: large memory requirement and slowness. In order to find the plane parameters accurately, parameter space must be divided finely in all three directions, and an accumulator assigned to each block. The Fast Hough Transform (FHT) described in [7] gives considerable speed up and reduces storage requirement.

The FHT applies to those Hough transform problems in which the equation relating features to parameters is linear in the parameters. In this case each feature votes for a *hyperplane* in parameter space (a $k-1$ dimensional generalisation of a plane, where k is the dimension of parameter space). The parameters are scaled so that their initial ranges form a 'hypercube' (generalisation of a cube) in parameter space.

A coarse Hough Transform is applied to the initial 'root' hypercube in parameter space by dividing it into 2^k 'child' hypercubes formed by halving the root along each of the k dimensions and assigning an accumulator to each child. Each hyperplane passing through a child hypercube contributes a vote to its accumulator. (In fact, a hyperplane is tested for intersection with a hypercube's circumscribing 'hypersphere'. This is approximate but is faster than the exact method.) Those children receiving greater than a threshold T votes are recursively subdivided, and so on. A limit is set on the level of subdivision, which is equivalent to setting a required accuracy.

An extra speed up is possible by keeping track of which features vote for (i.e. which hyperplanes intersect) each hypercube. Only those features need be tested for intersection between hyperplane and child hypercubes, since children lie inside their parents.

Plane finding using the FHT

In the FHT plane finder, 'hyperplanes' are planes and 'hypercubes' are cubes. The initial range of the parameters are: a: is -0.6 to 0.6. b: -0.8 to 0.8. c: $d_{peak} - 0.4H$ to $d_{peak} + 0.4H$ where d_{peak} is the disparity of the peak in the disparity histogram. The vertical disparity gradient limit b is set larger than the horizontal limit a because Needles is less sensitive to the shear distortion between images caused by b than the horizontal compression/expansion caused by a, since a shear preserves the area of an image patch.

The FHT threshold T is set to $0.6W$. A lower threshold would allow a fit to a smaller number of points, but would slow the algorithm down. The value $0.6W$ has been suitable for all the stereopairs so far tested. Normalising T using W is justified since one edge can only contribute one vote to the winning hypercube (this is proved in [8]), so that the value of T used implies that at least 60% of the edges in a left image patch must be matched.

Right Image Texture Density Test

The FHT plane fitter implicitly maps the left image patch into a parallelogram-shaped patch of the right image specified by the plane parameters (a, b, c). If the two patches are indeed projections of the same surface in the world, the texture density of the two patches should be similar. This test rejects a planar fit if the right image patch contains disproportionally more edges than the left image patch (in the reverse case a good planar fit could not have been obtained in the first place). Thus the planar fit is rejected if the ratio of number of right image patch edges to left image patch edges is greater than a threshold, set to $1.5(1 + a)$. $1 + a$ is the ratio of areas of the patches (right/left), so the test therefore rejects the planar fit if the number of right image edges is more than 1.5 times what we would expect.

1.4 Region Formation and Hypothesis Disambiguation

The next step is to join surface hypotheses in adjacent image patches if the two surfaces agree in their area of overlap. The test is based on the number of common matches in the sets of matches selected by the FHT. If this is ≥ 5, the hypotheses are labelled as connected. When one hypothesis could be connected to several hypotheses in the same patch, the one with the best planar fit is chosen, i.e. the one reaching the highest subdivision level in the FHT, or failing that the FHT accumulator values are compared.

A region growing process now explicitly labels regions of connected hypotheses. Each hypothesis becomes part of a numbered region. Hypothesis disambiguation then eliminates all but one of the surface hypotheses in each patch, with the set of remaining hypotheses assumed to be true representations of the visual surface. Discontinuities are located implicitly at the boundaries of regions. The stages are interleaved in the following way:

1. **First region growing stage.** Regions of connected hypotheses are grown by taking those hypotheses that are connected to neighbours in all eight (lateral and diagonal) directions as 'seeds', which grow into the network of hypotheses along their connections. Such hypotheses are used in descending order of the goodness of the planar fits. A seeded region expands breadth-first along the eight connection directions.

2. **Hypothesis disambiguation.** Competition between hypotheses in an image patch is resolved according to the 'strength' S of a region, calculated by summing the FHT subdivision levels of all the hypotheses in the region. Hence region strength represents the area of the region and the strength of the hypotheses within it. In each patch, the surface hypothesis belonging to the region of highest strength is declared the winning hypothesis in the patch. Hypotheses not part of any region are rejected.

3. **Second region growing stage.** Connections to hypotheses eliminated at the previous step are removed, and all region data is nullified. The region growing step 1 is then repeated. This is necessary because step 2 may split a region into two parts, still wrongly labelled as the same region.

4. **Elimination of weak regions.** The strengths of all the regions are recalculated. Those regions whose strength S falls below a threshold (20) are removed. This is designed to eliminate only very small regions.

Each continuous textured surface in the scene should be represented as a single region. Boundaries of a region should correspond to boundaries of the surface, e.g. step discontinuities, object boundaries. Note that a region may contain a step discontinuity if a connection route exists around it.

1.5 Least Squares Plane Fitting

The final stage in the Needles algorithm is to obtain more precise estimates of the local plane surface parameters than the quantised values obtained from the FHT plane finder. For the winning hypothesis in each patch, orthogonal regression is used to fit a plane to the disparity points that contributed to the winning plane in the FHT, minimising the sum of the squared perpendicular distances of the disparity points from the plane. Mathematical details are given in [13].

2 Parallel Implementation

The most time consuming parts of Needles take place independently in each image patch. The only non-local steps are region growing and final disambiguation, which take very little time. There is therefore great scope for using parallel processing to increase efficiency. Needles has been implemented on the MARVIN Transputer architecture developed in AIVRU, which is described in [3]. Using a nine Transputer system gives approximately an eight-fold decrease in running time over a Sun 3/60. Since one processor could in theory be assigned to each image patch, this is clearly a limited parallel implementation of Needles.

3 Results

We present results for two steropairs. The first is of a human face model, shown in figure 2. The face has been fixed to a backplate, painted white and dotted using a black pen to introduce texture. The images are 512×512 pixels, each pixel having a grey value between 0 and 255. Figure 3 shows the orientations of the local planar patches found by the Needles algorithm, shown superimposed on the left image. Each pin is centred on the centre of an image patch. The needles (hence the name) point in the direction of the surface normal in the world. The surface of the face is shown in figure 4a, plotted in world space. This was constructed by fitting a surface to the edge disparity points. For comparison, figure 4b shows the results of running the

PMF stereo algorithm [10]. This general-purpose algorithm employs a disparity gradient limit between matches that support each other, a weaker constraint than that imposed by Needles, resulting in occasional bad matches that cause glitches in the fitted surface. Since PMF is about two magnitudes faster than Needles, an obvious research direction is to try to incorporate the speed of PMF and the surface smoothness of Needles, to take advantage of both.

We have compared measurements of the face produced by Needles with height measurements of the face above the backplate made along cross-sections using a clock gauge. Both sets of measurements were relative to fixed axes marked on the backplate. The results are shown in graphs A to K of figure 5, representing the positions on the face shown in figure 4a. Solid squares mark the clock gauge data, outlined squares the Needles data. Gaps in the clock gauge data represent places where the slope was too steep for an accurate measurement to be made. Sub-millimetre accuracy has been achieved over large parts of most of the cross-sections, corresponding to sub-pixel accuracy in disparity. The large errors in graphs B, E and I seem to be caused by prominent surface features which are smoothed over by Needles, such as the eyes (E) and mouth (I).

The second stereopair, shown in figure 6, contains six more or less textured objects. The images are again 512×512 pixels. Figure 7 shows the planar surface normals found by Needles. Needles segmented the scene into the separate objects, in particular finding the discontinuity between the lego house and the telephone directory. The surfaces of the objects are shown in figure 8a. We also ran PMF and imposed the segmentation provided by Needles on the disparity data. The result of surface fitting is shown in figure 8b. The main difference between the results is the smoothness of the Needles surface for the telephone directory and the book. Repetitive texture, such as that on the book, is difficult for stereo algorithms to match correctly, since shifting either image by the 'wavelength' of the texture gives a match that is almost as good as the correct one. The global disambiguation mechanism employed by Needles causes the correct matches to be chosen since they will make up the largest region of connected surface patches.

4 Conclusion

We have implemented a stereo algorithm designed for smooth textured surfaces. It uses a local surface smoothness constraint and a novel global disambiguation mechanism that locates each surface as a whole. The algorithm has been implemented on a Sun and a network of Transputers. Extensions that have been made to the algorithm include crease discontinuity detection and calculation of surface curvature. These are described in [8].

5 Acknowledgements
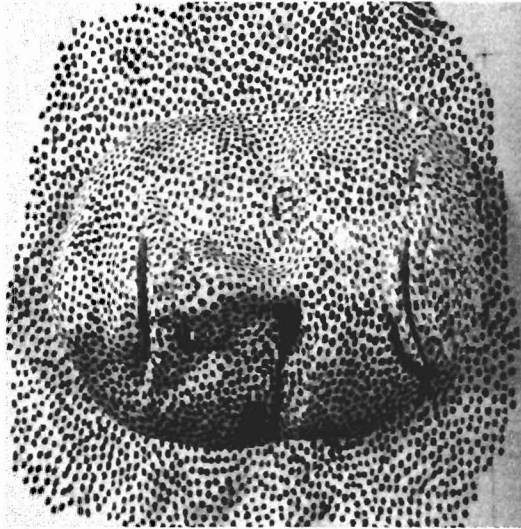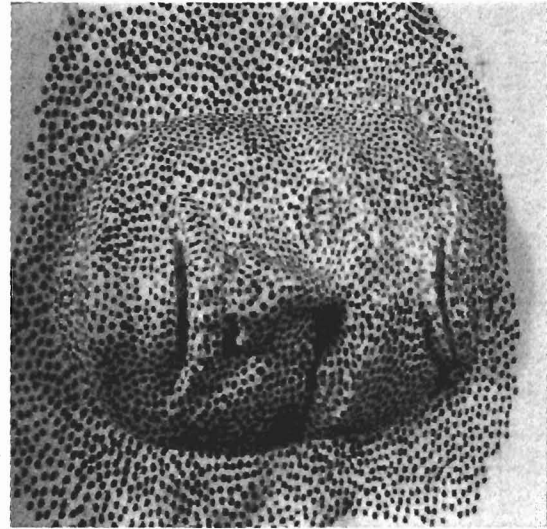
Right   Left
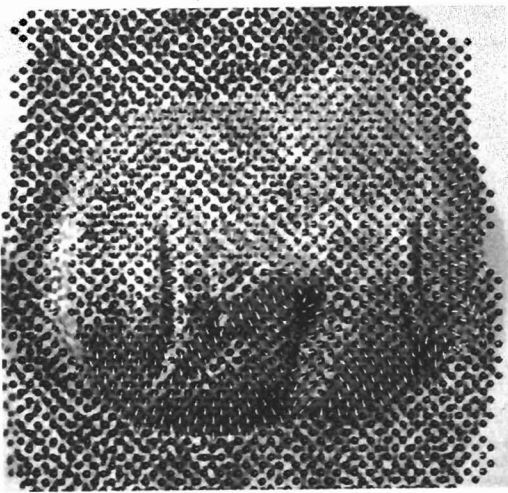
Figure 2: Stereopair of head model.

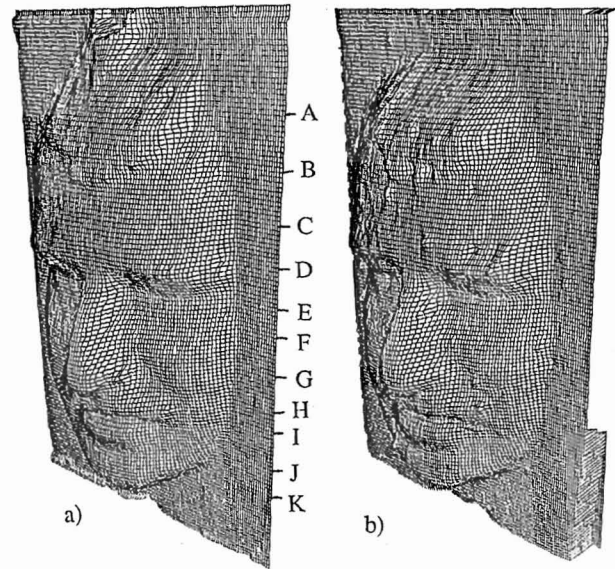Figure 3: Local planar surface normals found by Needles algorithm.

Figure 4: Surfaces of head found by a) Needles b) PMF.
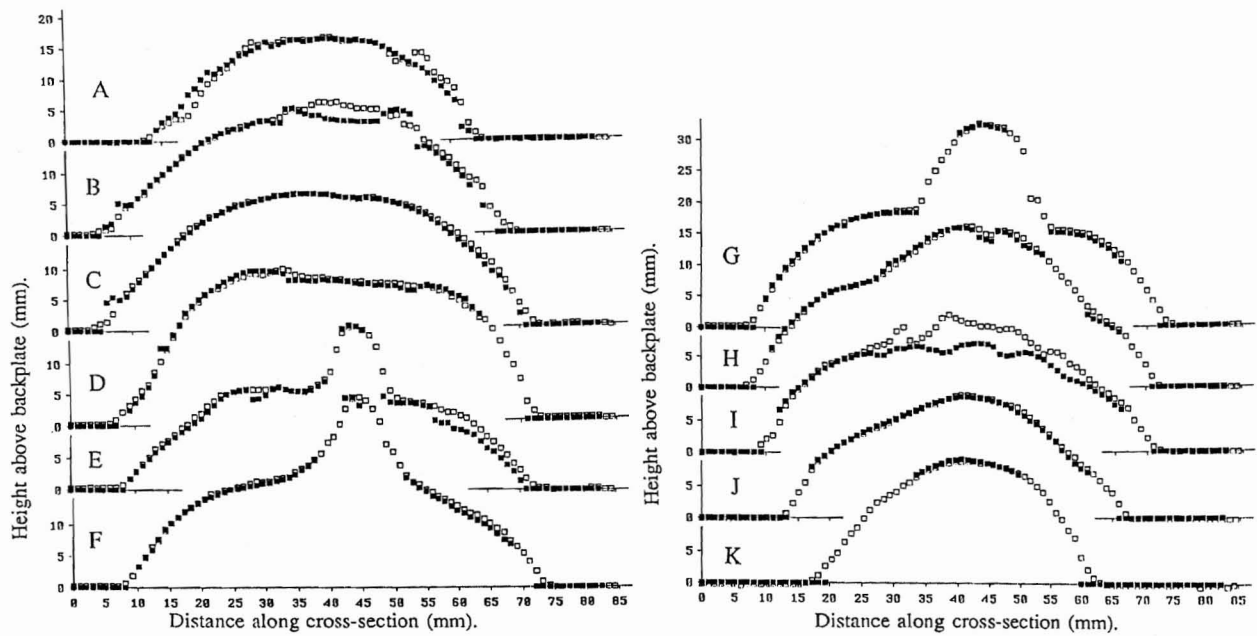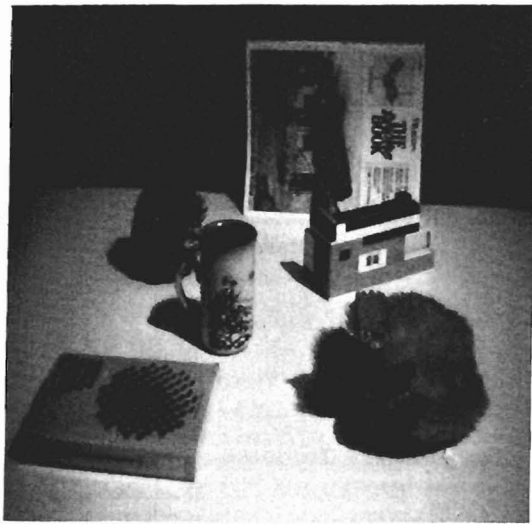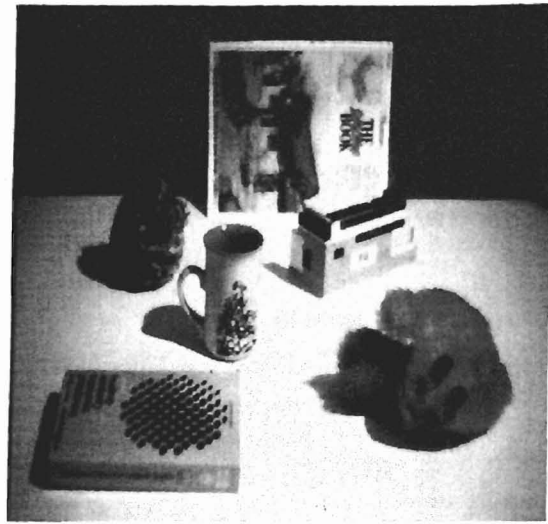

Figure 5: Cross-section graphs of head.

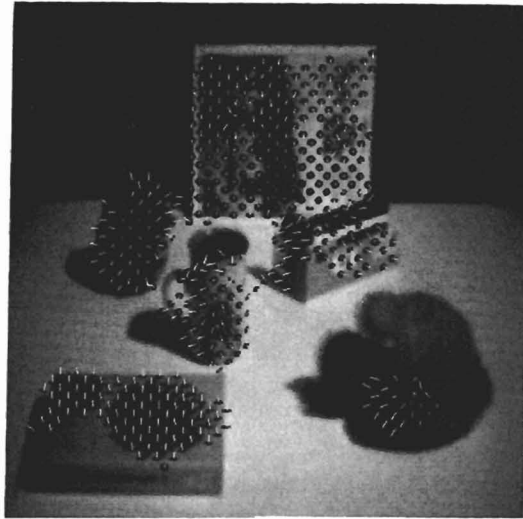Right　　　　　　　　　　　Left

Figure 6: Stereopair of six textured objects.

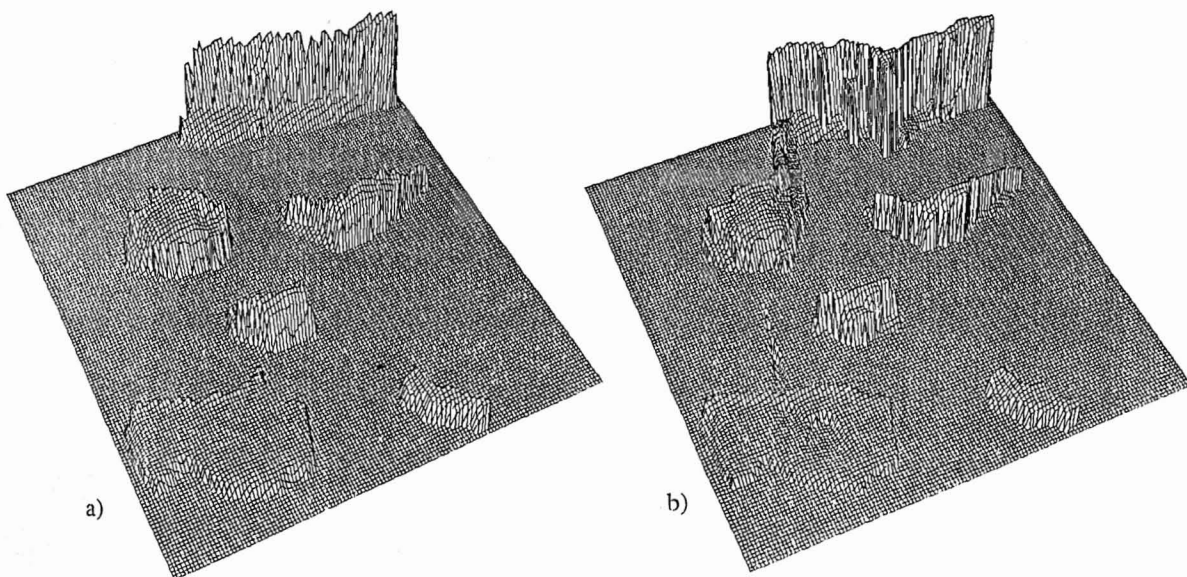Figure 7: Local planar surface normals found by Needles.

a)　　　　　　　　　　　b)

Figure 8: Surfaces of six objects, showing segmentation. a) Needles b) PMF, using segmentation provided by Needles.

References

[1] Barnard, S.T. and W.B. Thomson "Disparity Analysis of Images" *PAMI.* Vol. 2 No. 4 (1980) pp 333-340.

[2] Boult, T.E. and L. Chen "Synergistic Smooth Surface Stereo" *Proc. 2nd ICCV.* (1988) pp 118-122.

[3] Brown, C. and M. Rygol "An Environment for the Development of Large Applications in Parallel C" *Proc. Transputer Applications '90* (1990)

[4] Canny, J.F. "Finding Edges and Lines in Images" MSc thesis, MIT (1983).

[5] Grimson, W.E.L. *From Images to Surfaces: A Computational Study of the Human Early Visual System* MIT Press (1981).

[6] Hoff, W. and N. Ahuja "Surfaces from Stereo: Integrating Feature Matching, Disparity Estimation, and Contour Detection" *PAMI.* Vol. 11 No. 2 (1989) pp 121-136.

[7] Li, H., M.A. Lavin and R.J. Le Master "Fast Hough Transform: A Hierarchical Approach" *CVGIP.* Vol. 36 (1986) pp 139-161.

[8] McLauchlan, P.F. "Recovery of Textured Surfaces using Stereo Vision" PhD thesis. AIVRU, University of Sheffield (1990).

[9] Otto, G.P.and T.K.W. Chau "A 'Region-Growing' Algorithm for Matching of Terrain Images" *Proc. 4th AVC* (1988) pp 123-128.

[10] Pollard, S.B., J.E.W. Mayhew and J.P. Frisby "PMF: A Stereo Correspondence Algorithm using a Disparity Gradient Limit" *Perception.* Vol. 14 (1985) pp 449-470.

[11] Pollard, S.B., J.E.W. Mayhew and J.P. Frisby "Implementation Details of the PMF Stereo Algorithm" in this volume.

[12] Porrill, J., S.B. Pollard, T.P. Pridmore, J.B. Bowen, J.E.W. Mayhew and J.P. Frisby "TINA: The Sheffield AIVRU Vision System" *Proc. 10th IJ-CAI* Vol. 2 (1987) pp 1138-1144.

[13] Porrill, J., T.P. Pridmore, J.E.W. Mayhew and J.P. Frisby "Fitting Planes, Lines and Circles to Stereo Disparity Data" Technical Report 17. AIVRU, University of Sheffield (1986).

[14] Shirai, Y. and Y. Nishimoto "A Stereo Method using Disparity Histograms of Multi-Resolution Channels" *Proc. Robotics Research: 3rd Int. Symp.* (1986) pp 27-32.

[15] Terzopoulos, D. "Multiresolution Computation of Visible-Surface Representations" PhD thesis. MIT (1984).

[16] Tsai, R.Y. "An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision" *Proc. CVPR '86.* (1986) pp 364-374.