SWITCHER:
A stereo algorithm for Ground Plane Obstacle Detection

Y.Zheng[*], D.G.Jones[#], S.A.Billings[*], J.E.W.Mayhew[#], J.P.Frisby[#]

[#]Artificial Intelligence Vision Research Unit, University of Sheffield.
[*]Department of Control Engineering, University of Sheffield.
Sheffield, S10 2TN.

A stereo algorithm for detecting and locating obstacles raised above a ground plane is described. It is based on applying a simple multi-model approach using hypothesis tests to the grey levels of pre-processed stereo images. The algorithm has been successfully applied to real images captured from a moving vehicle in order to recover putative obstacle locations.

1. INTRODUCTION

We describe an algorithm, called SWITCHER, for analysing stereo images for the purposes of vehicle obstacle detection. The algorithm has been developed as part of a project[1] whose overall aim is the demonstration of a stereo vision system capable of warning the driver, human or automaton, of obstacles in the path of the vehicle that constitute a potential hazard given the current vehicle trajectory. It is assumed that the terrain over which the vehicle is travelling is a reasonably even surface, such as a path, road or factory floor.

In brief, SWITCHER finds disparity matches as follows:

(a) Each stereo image is first subjected to a stage of low level image pre-processing which includes: image rectification to parallel camera geometry without aliasing, high pass filtering, and grey scale equalisation.

(b) A window of grey levels from the left image is incrementally moved in 1 pixel shifts across the right image, subtracted from the right image at each location, and the variance of the differences so obtained is calculated for each disparity shift. Shifts are restricted to the range expected for obstacles protruding from the ground plane.

(c) A statistical test (the F-test) is used to decide which disparity shift gives the smallest expected variance in the differences produced by subtraction. This test is applied sequentially to points along rasters and a decision is made to switch from the disparity selected for the previously considered window location only if there is evidence of a significantly lower variance of the difference at a new disparity value: hence the name SWITCHER for this algorithm.

It is demonstrated using natural images collected from vehicle-mounted cameras that SWITCHER is capable of

delivering a disparity map from which the scene coordinates of a potential obstacle can be derived.

2. SPECIFICATION OF THE PROBLEM

The problem of detecting and locating obstacles raised above a ground plane by stereo vision may be studied within the following framework.

Defining any two matching image rasters

$$l(x) = s_l(x) + n_l(x)$$
$$r(x) = s_r(x) + n_r(x) \tag{1}$$

where x is the image coordinate of corresponding points on the ground plane, $l(x)$ and $r(x)$ are intensities recorded by the left and the right cameras respectively, and $n_l(x)$ and $n_r(x)$ are noise components presumed to arise from camera noise, image capture, image pre-processing, etc. When no obstacle is present, and taking s to be related to the photometric properties of points on scene surfaces, we make the following assumption

$$s_l(x) = s_r(x) \tag{2}$$

That is, for the purposes of analysis, we assume that the pre-processing stages are sufficient to eliminate any problems arising from the stereo projections failing to possess photometric invariance. Experiments to date show that this assumption is reasonable. Indeed we have obtained good results for several images sequences without image pre-processing other than rectification.

When an obstacle is present whose surface plane has a uniform disparity d, then

$$s_l(x) = s_r(x-d) \tag{3}$$

and the problem will be specified as: find d and determine the domain of x upon which eq.(3) holds.

3. SOME NOTATION AND ASSUMPTIONS

Initially, to simplify the description, assume that only one obstacle is present along a raster. The extension of the algorithm to several obstacles can easily be made. Also, assume the disparities of the obstacle and the ground plane are each uniform along a raster (this assumption is not restrictive given a reasonably small window size). The study will be carried out on a raster basis, i.e., the image signals $l(x)$ and $r(x)$ are both one-dimensional. The extension to two-dimensions is straightforward.

Fig.1 shows the simplest case in which only one obstacle with uniform disparity d across the raster is present. Δ is the width of the obstacle along the raster.
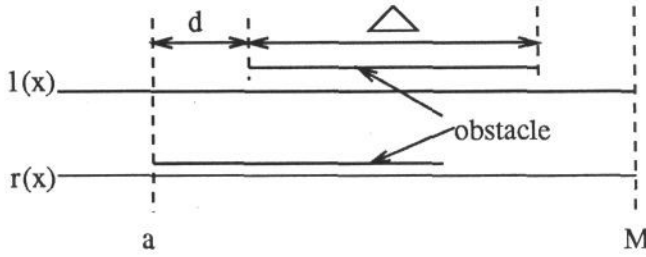

Fig. 1

For ease of reference in the following analysis, several definitions are introduced. Let $P_b(x)$ and $P_o(x)$ be the background and the obstacle image signals of the left raster respectively. Note that both of them are functions of the pixel position x. Let GP, OC and OB be defined as three regions of x with the following properties

$$GP = \{x/\ s_l(x) = P_b(x) \ \text{ and } \ s_r(x) = P_b(x)\} \qquad (4)$$

$$OC = \{x/\ s_l(x) = P_b(x) \ \text{ and } \ s_r(x) = P_o(x+d)\} \qquad (5)$$

$$OB = \{x/\ s_l(x) = P_o(x) \ \text{ and } \ s_r(x-d) = P_o(x)\} \qquad (6)$$

with $d > 0$.

Clearly GP represents the region of the ground plane which can be seen by both cameras; in Fig.1 this is the region where $1 \le x \le a$ and $a+d+\Delta+1 \le x \le M$. Region OC represents the occlusion part that can be seen by the left camera but not the right; in Fig.1 OC is the region where $a+1 \le x \le a+d$. The obstacle which can be seen by both cameras but with a forward phase shift of the right signal relative to that of the left is located in region OB; this region is given by $a+d+1 \le x \le a+d+\Delta$ in Fig.1.

Now define various differences as follows

$$e_0(x) = l(x) - r(x),$$
$$e_1(x) = l(x) - r(x-1),$$
$$\cdot$$
$$e_{d_m}(x) = l(x) - r(x-d_m)$$

where d_m is the maximum expected disparity of an obstacle, and the set e_0 to e_{d_m} can be regarded as a multi-model formulation.

4. OBSTACLE DETECTION

It is obvious from the above definitions that when no obstacle is present, i.e. for all $x \in GP$,

$$e_0(x) = n_l(x) - n_r(x),$$
$$e_1(x) = s_l(x) - s_r(x-1) + n_l(x) - n_r(x-1),$$
$$\cdot$$
$$e_{d_m}(x) = s_l(x) - s_r(x-d_m) + n_l(x) - n_r(x-d_m)$$

Assume the noise signals $n_l(x)$ and $n_r(x)$ are normally and independently distributed with zero mean and variance σ_n^2 and assume they are uncorrelated with s_l and s_r. Then the zeroth difference $e_0(x)$ is also normally distributed with

zero mean and variance $2\sigma_n^2$. However, the variances of the rest of the differences $e_1(x), \ldots, e_{d_m}(x)$ are all expected to be higher than $2\sigma_n^2$ unless the ground plane has no texture. Similarly, for all $x \in OB$, the differences are given by

$$e_0(x) = s_l(x) - s_r(x) + n_l(x) - n_r(x),$$
$$\cdot$$
$$e_d(x) = n_l(x) - n_r(x-d),$$
$$\cdot$$
$$e_{d_m}(x) = s_l(x) - s_r(x-d_m) + n_l(x) - n_r(x-d_m)$$

and $e_d(x)$ becomes normally distributed with zero mean and variance $2\sigma_n^2$. The variances of the rest of the differences are expected to be higher than $2\sigma_n^2$ unless the obstacle surface has no texture. The above observation suggests the use of the variances of the differences as the key to locate an obstacle and to determine its disparity.

A simple method to detect the position of an obstacle would be to calculate the zeroth difference $e_0(x)$ only and then set up a threshold such that if in some domain Ω, $e_0(x)$ is above the threshold for almost all $x \in \Omega$, an obstacle may be present in Ω. This is the basic idea used by Mallot et al [1] in their obstacle detection method. It is quick to implement, but does not provide information on the disparity of the obstacle. Also, problems arise in setting a threshold suitable for the particular image textures presented by the scene under consideration. It was this latter difficulty that led us to design SWITCHER, which avoids the problem of setting thresholds by using a statistical test. We are however continuing to investigate Mallot et al's approach, not least because of its simplicity and low computational cost [4].

5. THE NECESSITY FOR A STATISTICAL TEST

The method introduced here can recover disparity information. The idea is to set up a window of size W on differences $e_i(x)$, $i=0,1,\ldots,d_m$, estimate the variance of the differences in the window and then make comparisons. The window size W is taken as an odd number for convenience, e.g., the kth window on the zeroth difference is the set

$$\{\ e_0(x),\ x = k-\frac{W-1}{2},\ldots,k,\ldots,k+\frac{W-1}{2}\ \}$$

Let $\varepsilon_i(x)$ be the ith difference after the mean has been removed. The variance of this difference in the kth window can then be estimated by

$$\hat{\sigma}_i^2(k,\ W) = \frac{\displaystyle\sum_{x=k-(W-1)/2}^{k+(W-1)/2} \varepsilon_i^2(x)}{W-1} \qquad (7)$$

where

$$i=0,1,\ldots,d_m$$

As noted above, provided all $x \in GP$ in the kth window, $\hat{\sigma}_0^2(k,W)$ has the minimum expected value $2\sigma_n^2$. Similarly, provided all $x \in OB$ in the kth window, $\hat{\sigma}_d^2(k,W)$ has the minimum expected value $2\sigma_n^2$. Thus the exercise of locat-

ing an obstacle is reduced to finding the minimum expected value of $\hat{\sigma}_i^2(k,W)$, $i=0,1,...,d_m$ at every window. The disparity d at the kth pixel of the left image will then be given by

$$E[\ \hat{\sigma}_d^2(k,W)\ ] = \min_i E[\ \hat{\sigma}_i^2(k,W)\ ] \qquad (8)$$

where $E[z]$ denotes the expected value of the random variable z. However, when some x in the kth window lies in the region OC, or some $x \in OB$ and some $x \in GP$, the kth window enters the occlusion parts and all variances of the differences are expected to be greater than the minimum $2\sigma_n^2$.

The basic problem now is to solve eq.(8). Suppose in the $(k-1)^{th}$ window, $d=0$; in the kth window, it is found that $\hat{\sigma}_p^2 < \hat{\sigma}_0^2$ for some $i=p \neq 0$. A decision has to be made here on whether $\hat{\sigma}_p^2 < \hat{\sigma}_0^2$ by chance or $\hat{\sigma}_p^2$ actually comes from a population with expectation of mean less than that of $\hat{\sigma}_0^2$. The above situation clearly calls for a statistical test. If, after the statistical test, it is found that $\hat{\sigma}_p^2$ is not significantly smaller than $\hat{\sigma}_0^2$, then the disparity remains at zero. Otherwise the decision $d=p$ can be made. The statistical test used is the F-test where, in the above example,

$$F = \frac{\hat{\sigma}_0^2(k,W)}{\hat{\sigma}_p^2(k,W)}$$

if certain conditions are met by $l(x)$ and $r(x)$.

6. APPLICATION OF THE F-TEST TO OBSTACLE DETECTION

Consider the case when $x \in GP$. (Similar arguments hold in the case $x \in OB$.) Then,

$$e_0(x) = n_l(x) - n_r(x)$$
$$e_i(x) = s_l(x) - s_r(x-i) + n_l(x) - n_r(x-i)$$

where

$$i = 1,...,d_m$$

To use the F-table to compare $\hat{\sigma}_0^2$ and $\hat{\sigma}_p^2$ for some $i=p \neq 0$, it is necessary that both e_0 and e_p are from two independent populations with normal distributions. The above two equations show that this requirement cannot be satisfied completely. To see this, assume both s_l and s_r are normally distributed with the same variance σ_s^2. (On the whole, this assumption will be valid.) Then the two differences e_0 and e_p are correlated due to the noise signal n_l, and their correlation is

$$\phi_{0p}(0) = \frac{\sigma_n}{\sqrt{\sigma_n^2 + \sigma_s^2 + R(p)/2}}$$

where

$$R(p) = E[\ (s_l(x)-\bar{s}_l)(s_r(x-p)-\bar{s}_r)\ ]$$

$\bar{s}_l$ and $\bar{s}_r$ are the means of s_l and s_r respectively. Thus to reduce their correlation, the image must have a sufficiently large signal-to-noise ratio[2]

$$\frac{S}{N} = \frac{\sigma_s}{\sigma_n}$$

The second point to note in applying the F-test is that the F-table is produced on the basis of two populations having the same variance. Hence in deciding if $\hat{\sigma}_p^2$ is significantly smaller than $\hat{\sigma}_0^2$, the hypothesis to be tested is that both of them have the same expectation. When this hypothesis is accepted, it seems that the disparity could be either 0 or p. At this point, the decision made in the $(k-1)^{th}$ window will influence the present decision in the kth window, the argument being that a decision to change disparity must be supported by adequate evidence. Although such a criterion makes the change of disparity a damped process, it reduces noise in the disparity estimation. Indeed, in order to avoid too hasty switching to a new disparity we have explored switching only if three successive F-tests are all significant. That procedure can be effective in avoiding false alarms but it is not always necessary and its potential is the subject of current experimentation [4].

The third point is that the decision on whether $\hat{\sigma}_p^2$ is significantly smaller than $\hat{\sigma}_0^2$ is a one-sided test on the F-distribution.

Thus, suppose in the $(k-1)^{th}$ window it was found that $d=q$ (i.e. q is the presently chosen disparity value). Using H for hypothesis and denoting the level of significance by α, the steps taken to test the hypothesis of $d=q$ in the kth window, which together define the SWITCHER algorithm, are as follows.

1. H: $d=q$.

2. Choose α.

3. Calculate $\hat{\sigma}_i^2(k, W)$, $i = 0,1,...,d_m$.

4. Find p such that $\hat{\sigma}_p^2 = \min_i \hat{\sigma}_i^2$.

5. If $p = q$, accept H.

6. If $p \neq q$, find $F = \dfrac{\hat{\sigma}_q^2}{\hat{\sigma}_p^2}$.

7. Reject H if $F > F_{1-\alpha}(W-1,W-1)$ and set $d=p$.

Noise on the disparity estimation will be introduced for the occlusion parts as none of the differences will have the minimum expected variance. The occlusion parts are transition periods between different disparities. Provided the image has sufficiently large signal-to-noise ratio and provided the window size is properly chosen, false estimates of disparity for the occlusion parts should not result in false disparity estimates when $x \in GP$ or $x \in OB$.

6.1. Two-dimensional windows

Instead of working on a single raster, the above study can easily be extended to include several rasters provided an obstacle present on these rasters has a reasonably uniform disparity from one raster to its neighbouring rasters. In fact, in the demonstrations which follow matching windows straddling three rasters are used.

We have examined both theoretically and experimentally the effects of varying the window size (Zheng et al, in

---

[2] Experiments using artificial random dot textures have shown that S/N ratios at least as low as 5:1 are satisfactory, and experience with natural images suggests that even lower values can be tolerated.

preparation). Our main conclusion is that the smallest obstacle that can be reliably detected is about half the width of the window but that vulnerability to noise increases as window size is reduced. The choice of 17×3 provides a good compromise for the image sequences investigated to date.

7. IMPROVEMENTS TO PERFORMANCE

7.1. The Spatial Coherence Constraint

Some of the road scenes we have investigated have very low S/N ratios caused by poor dynamic range in the TV camera response when dealing with low intensity regions within otherwise brightly lit scenes. We have found that the problem of the false alarms triggered in such regions can be met in large part by employing a *spatial coherence constraint*. The idea here is that certain 'noise spikes' in the disparity map could not arise from the scene without having given rise to above-ground plane disparities recorded 'behind' them *unless* the cameras happened to be looking at a thin lamina so oriented that it is seen 'edge on'. It is safe to assume that such circumstances are unlikely. Hence the absence of supporting disparities above the ground plane gives a firm basis for excluding these noise spikes. We have found this constraint a highly effective noise cleaner.

7.2. Exploiting the ground plane disparity

We have found it advantageous both in speed and in noise elimination to limit the range of disparties that is searched to about + 40 pixels. This range is ample to capture obstacles in the images explored to date (but see remarks in section 10 on exploiting temporal coherence between successive stereo pairs).

8. EXPERIMENTS USING SWITCHER ON NATURAL IMAGES

SWITCHER has been evaluated off-line using a selection of 256×256 stereo image pairs recorded on video from cameras mounted $52\ cm$ apart and $67\ cm$ above the ground on the RSRE experimental vehicle.

8.1. Image Pre-processing

Each image pair was first subjected to a series of geometrical and grey level transformations. The aim of this image pre-processing was to recover the underlying texture within each raster in a form most suited to the simple grey level subtraction operation upon which SWITCHER is founded.

First, a point transform rectification with anti-aliasing by local neighbourhood Gaussian smoothing to sub-pixel acuity was applied to the images to achieve zero vertical disparity between corresponding left and right rasters. Knowledge of the camera geometry required for this stage was provided in the present experiments by an implementation of Tsai's algorithm [3] in AIVRU's *TINATOOL* image processing environment [5].

Second, local variations in grey levels caused by photometric variance between images, low frequency drifts, etc were then reduced to acceptable levels by high-pass filtering with a low cut-off frequency of 2 cycles/deg.

Finally, a two band log grey scale equalisation filter, with each band symmetrically placed around the mean grey level, was applied to enhance homogeneity of textural information in the signals by emphasising their low amplitude portions.

The last two stages are not always necessary for SWITCHER to perform adequately. Experiments are in progress to investigate when they are required [4].

8.2. Variance Residue Measurements & Disparity Selection

vs 11 A 17×3 window from the left image was incrementally moved in 1 pixel steps across the right image, subtracted from the right image at each location, and the variance calculated for · the differences so obtained. A selected disparity value for each location was obtained by comparing variances using the F-test described above with a 5% significance level throughout.

9. RESULTS

The results from applying SWITCHER to two sample stereo pairs are shown in figures 2 and 3. The rectified stereo images are shown first (a - right image, b - left image, as required for crossed-eye fusion), followed by the results of high pass filtering and grey scale equalisation of the left image, (c). The disparity map returned by SWITCHER is shown in (d), and (e) illustrates the result of imposing on that map the constraint of spatial coherence.

9.1. The Post Image Pair

In figure 2e the disparity map returned for a region where the ground plane has no obstacle is clearly distinguishable from a region where the post-like obstacle is present. The smooth slanted obstacle disparity surface is bounded on each side by noisy regions caused by occlusions. It has proved a simple matter to estimate the x,y,z scene coordinates of the obstacle from this map.

Comparison of the equation for the recovered ground plane with that provided from the same images by a fast Hough Transform disparity mapping technique [2], shows that SWITCHER provides accurate disparity information.

9.2. The Kerb Image Pair

The right hand portion of the disparity map (3e) derives from the ground plane and is again reasonably flat, whereas the left hand portion from the kerb region is noisy. The latter is probably due to differences between the image projections of the high frequency grassy texture above the kerb, and a choice of window size unsuited to the rapid changes in disparity in the grassy area. Even so, SWITCHER's output is quite sufficient for the purposes of obstacle detection assuming a knowledge of the ground plane.
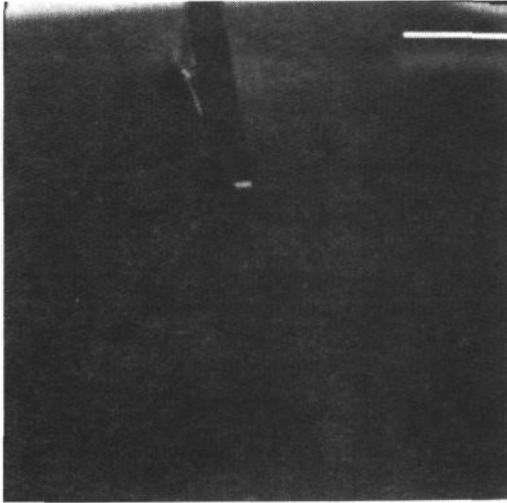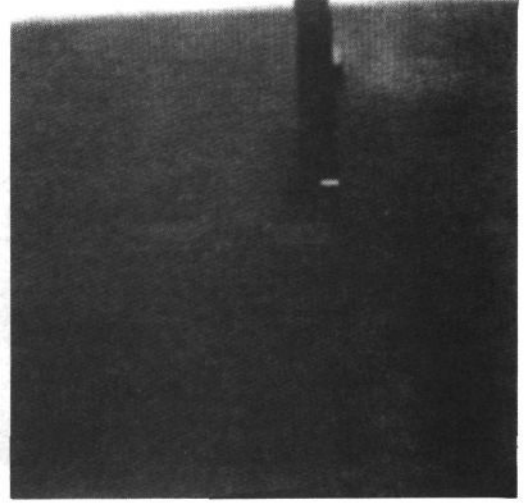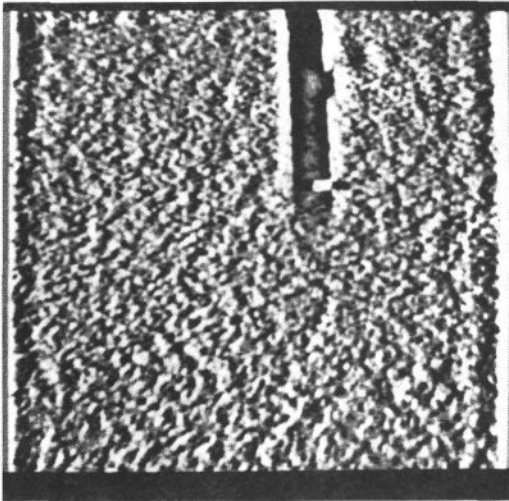
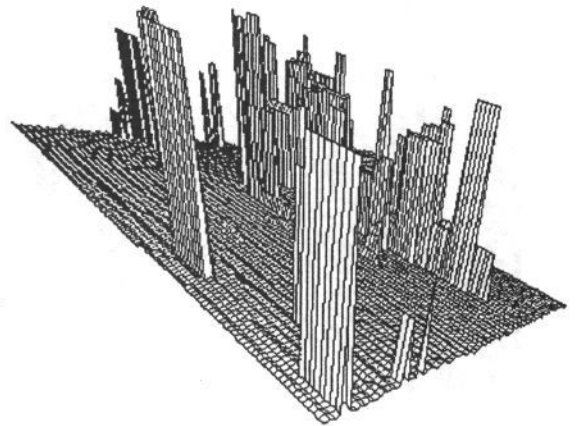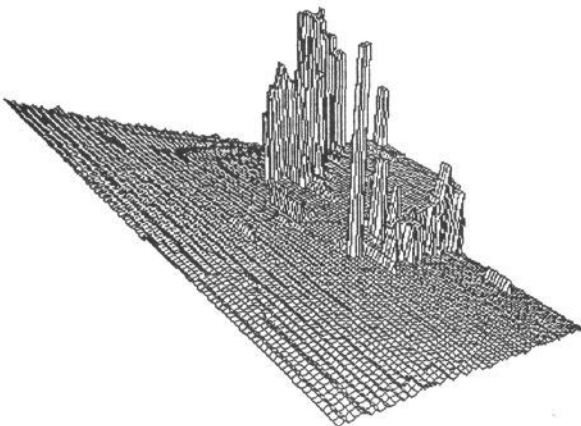Fig 2 (a)

Fig 2 (b)

Fig 2 (c)

Fig 2 (d)

Fig 2 (e)

10. CONCLUSIONS

The SWITCHER algorithm has been shown to be capable of delivering information regarding obstacles from real stereo images captured from a moving vehicle. In experiments to date, in no case is an obstacle missed and problems posed by false alarms appear slight [4]. Work is in hand to investigate versions of SWITCHER which preserve these attributes while reducing as far as possible the computational cost.

Information about a putative obstacle, delivered as a rough estimate of the x,y,z scene coordinates of the obstacle, could be used to control a region of interest processing strategy aimed at checking in more detail the image zone where the obstacle is recorded. Work is planned to use SWITCHER in this way within a control architecture that integrates information about the scene from a continuous stream of stereo image pairs collected on a moving vehicle, to provide advance warning about obstacles likely to cause a collision given the current vehicle trajectory.
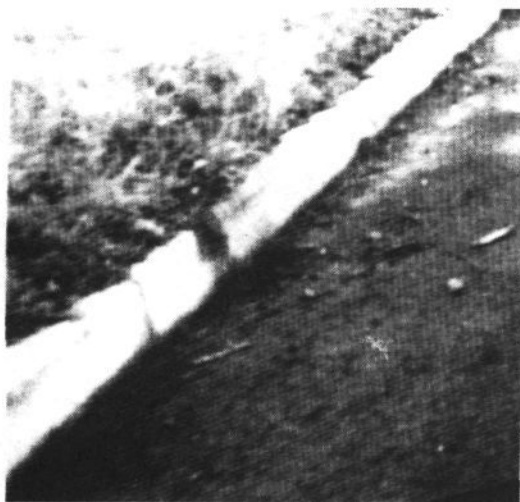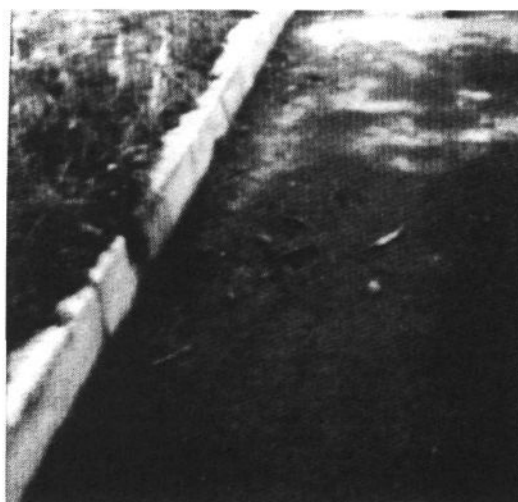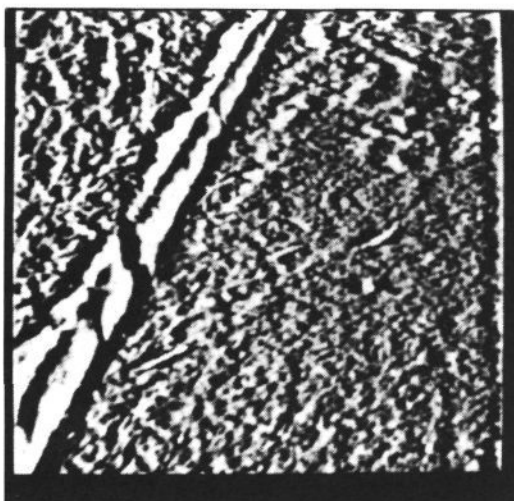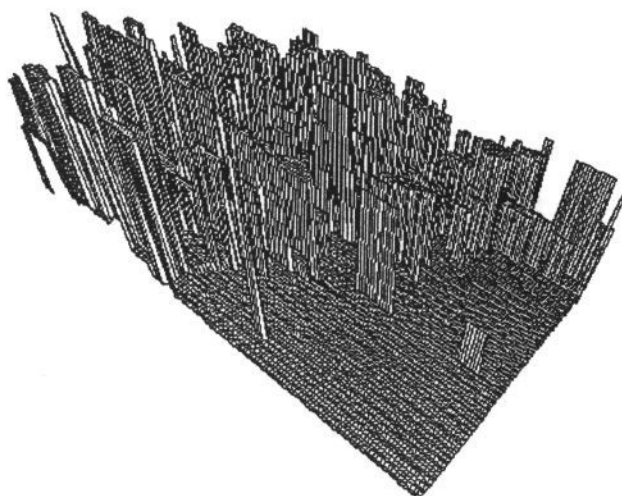
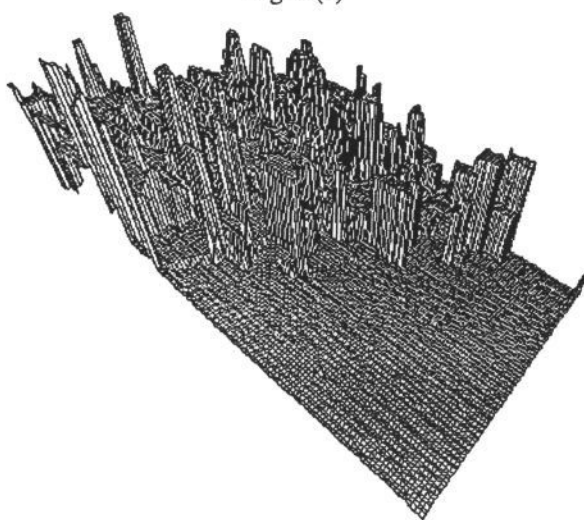Fig 3 (a)

Fig 3 (b)

Fig 3 (c)

Fig 3 (d)

Fig 3 (e)

11. REFERENCES

1. **H. A. Mallot, E. Schulze and K. Storjohann,** 'Neural Network Strategies for Robot Navigation', in G. Dreyfus and L. Personnaz (eds.) *Proc. nEuro '88* (Paris, June 1988).

2. **P. McLauchlan and J. E. W. Mayhew,** 'Needles: A Stereo Algorithm for Texture', *Workshop on Image Understanding and Machine Vision, Cape Cod Massachusetts* June 1989.

3. **R. Y. Tsai,** 'An efficient and accurate camera calibration technique for 3D machine vision', *Proc IEEE CVPR 86* 1986, pp 364-374.

4. **D. G. Jones, Y. Zheng, S. A. Billings, J. E. W. Mayhew and J. P. Frisby,** 'Experiments on stereo algorithms for ground plane obstacle detection', *In preparation.*

5. **J. Porrill, S. B. Pollard, T. P. Pridmore, J. B. Bowen, J. E. W. Mayhew and J. P. Frisby,** 'TINA: The Sheffield Vision System', *Proceedings of the Ninth International Joint Conference on Artificial Intelligence, Milan* 1987, pp 1138-1144.