

Ground Plane Obstacle Detection of Stereo Vision Under Variable Camera Geometry Using Neural Nets

Y. Shao, J. E. W. Mayhew, S. D. Hippisley-Cox
Artificial Intelligence Vision Research Unit
University of Sheffield
Psychology Building, Western Bank
Sheffield, S10 2TP
United Kingdom
yuan@aivru.sheffield.ac.uk

Abstract

We use a stereo disparity predictor, implemented as layered neural nets in the PILUT architecture, to encode the disparity flow field for the ground plane at various viewing positions over the work space. A deviation of disparity, computed using a correspondence algorithm, from its prediction may then indicate a potential obstacle. A casual bayes net model is used to estimate the probability that a point of interest lies on the ground plane.

1 Introduction

Being able to detect floor obstacles or deviations from planarity of the ground plane is essential for mobile robot navigation. The work reported here uses a four-degree-of-freedom stereo camera rig¹ to detect any obstacle lying on the ground plane, and is part of an ongoing project of 3-D stereo vision reconstruction and mobile vehicle control.

The most straightforward solution to the problem perhaps is to recover the depth of the whole scene, or a selected region of interest, by triangulation, then to fit this depth information into a plane model. This solution, however, requires to know the extrinsic parameters of both cameras to a high degree of accuracy. Unfortunately this is not usually the case. When the rig moves, even under control, a great uncertainty of camera position and pose may be present due to various sources of error: mechanical, dynamic, kinematic and control as well.

Therefore, statistical and probability-based theories are widely employed to work out various proposed solutions.

¹The stereo camera rig used for this work comprises two 3-link kinematic chains with rotation around Pan, Tilt, left and right Verges.

2 Previous Work

D. M. Booth, et al [2] in their paper present a two pass algorithm to detect ground plane obstacles using fixed cameras. First, they solve the correspondence problem by matching grey level intensity rasters. In the second pass, confidence measures of disparity estimations are computed by comparing the noise model with the residual error distribution of correctly matched rasters. The Bhattacharyya distance between two univariate distributions is used to construct an approximation of error probability, which gives a measure of belief that a given pixel represents a point that belongs to the ground plane.

M. R. M. Jenkin and A. Jepson [5] try to detect floor anomalies by verifying the planarity assumption. First a known 3D calibration object is used to calibrate the fixed cameras. With the resulting matrices and an initial rough estimation for the floor parameters, they are able to model the ground. Then the coefficients in this model are refined. Based on the constancy assumption, a phase-based disparity scheme is chosen to acquire the measurement of the disparity. Also a mixture model is employed to represent the disparity as arising from one of several simple distributions, specifically, those for outlier, for floor and for objects near floor. Finally, a so-called EM-algorithm (Estimation and Maximization) is constructed to compute "ownership likelihoods" at each pixel.

To detect ground plane obstacles under variable camera geometry, S. Cornell, et al [3] use a Parametrised Interpolated Look-Up Table (PILUT) [6], implemented as layered neural nets [10], to predict the features correspondence (disparity), which is indexed by head state. Any deviation from the predicted images coordinates then may indicate a point in space which is not on the ground plane. The experiment showed the successful discernment of a 2cm high obstacle at a distance of about 1m.

3 Our Approach

The work reported here is an extension of that done by S. Cornell, et al. We are not going to detect obstacles directly. Instead, we try to check whether a point of interest lies on the ground plane. To do so, we need to build a disparity model for the ground plane. Here, PILUT again is used to encode the disparity map of the ground plane so as to avoid the dependency upon the camera calibration and the dynamics of the camera rig. A modified Förstner corner detector is employed to find points of interest. A casual Bayes net is built to combine the disparity deviation of matching from prediction with the Bhattacharyya distance, which is used to measure the distance between the two residual error distributions. A likelihood measurement can then be derived from this Bayes net. The system is illustrated in Figure 1.

4 Encoding Disparity Mapping Using PILUT

The general principle of the PILUT is to use a linear combination of basis functions to approximate a multi-dimensional function. The tensor product of polynomial

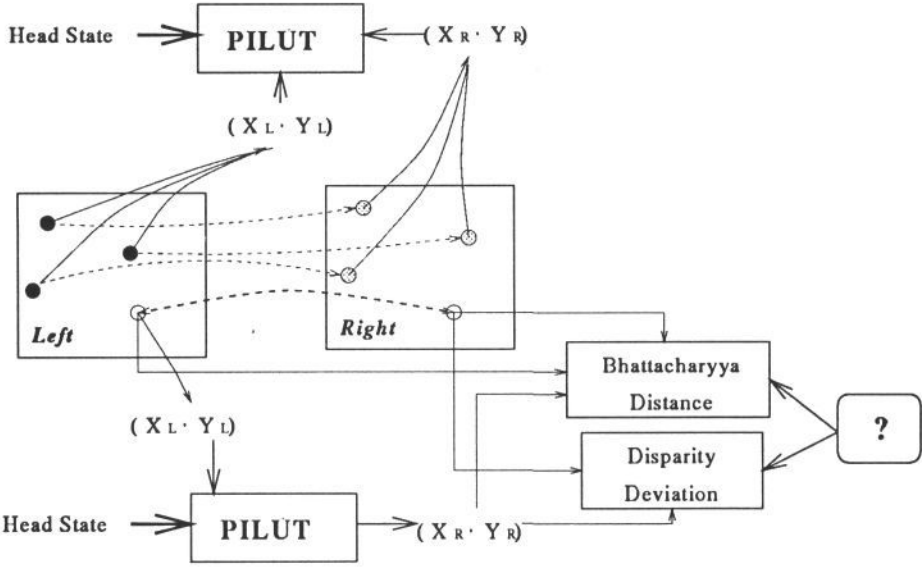


Figure 1: System scheme

and radial function bases is used to form the blended polynomial expansion.

For instance, to approximate

$$y = f(x_0, x_1, \dots, x_n),$$

we use a blended polynomial expansion of the form

$$y = \sum u_{ij} \phi_i(x) \psi_j(x),$$

where the $\phi_i(x)$ form a polynomial basis, (e.g. constant basis (1), affine basis $(1, x_0, x_1, \dots, x_n)$, or higher order basis, and the $\psi_j(x)$ are a radial basis (RBF), with

$$\psi_j(x) = g\left(-\frac{1}{2}(x - m_j)^T S_j^{-1}(x - m_j)\right),$$

where the m_j is the mean and the S_j covariance. To obtain gaussian RBF, we set $g = \exp$. After we have chosen our bases, the coefficients u_{ij} , m_j and S_j are then learned from training data using a scalar measurement kalman filter.

We use PILUT to approximate the functional relationship between disparity flow field and rig head state. There are 7 degrees of freedom (the pan of the rig being kept still during experiment): head tilt (T), left and right verge (LV and RV), x and y coordinates of a point of interest in left image (x_l and y_l) and in right one (x_r and y_r). The first five can be considered as inputs of the function while the last two are outputs.

Images captured are rectified using framerate hardware to approximately (pseudo) parallel camera geometry [9] will reduce the vertical disparity by a great amount.

However, for some bad camera geometries, they are still of the order of 10 pixels, as can be seen from the disparity map of Figure 2(1a).

Before PILUT is capable of predicting feature correspondence, we train the nets off-line with a set of training data. Training data is obtained by keeping the camera foveating a target on the ground floor [7]. A pair of images of a calibration tile is made so that we can use an earlier developed automatic calibration grid detector based on Förstner corner detection to provide 64 pairs of precisely corresponding corners at each head state.

When training PILUT, some parameters, like number of nodes, polynomial order, RBF, etc., need to be chosen. Generally speaking, the more the nodes and the higher the order, the better the fitting, but at the price of storage and computing time. To find suitable PILUT learning parameters, results by different PILUTs are compared. RMS (Root of Mean Squared) error and MAX error are used to evaluate PILUTs' performance. Experiment tells a PILUT with quadratic or affine polynomial basis and Gaussian RBF is competent to learn the disparity map. The resulting RMS error is .55 pixel and the MAX one 3.17 pixels for a typical head state. Figure 2 illustrates how well a trained PILUT predicts the disparity. Figure 3 gives PILUT's performance by pre- and post-learning comparison of frequency and cumulative distributions of retinal errors for a stereo image pair of a practical scene.

5 Correspondence

We check points of interest in an image so that we do not need to compute at every pixel. Corners, here detected by a modified form of the Förstner corner detection algorithm [4]², are chosen to represent the points of interest.

We refer to the coordinates of the i -th pixel and its gradient by p_i and f_i respectively, then the coordinates of a corner candidate are given by

$$p = \left(\sum_{i=1}^m W_i \right)^{-1} \left(\sum_{i=1}^m W_i \cdot p_i \right),$$

where

m : the number of pixels in the window,

$W_i = f_i f_i^T$, "square of gradient".

We use the following correlation algorithm to match a corner in the left image with one in the right image. Labelling the disparity at pixel (x, y) in the left image by (d_x, d_y) , we set

$$\hat{\sigma}_{d_x, d_y}(x, y) = \min_{i, j} \hat{\sigma}_{i, j}(x, y),$$

where

$$\hat{\sigma}_{i, j}^2(x, y) = \frac{1}{W} \sum_{i, j} W_{i, j} (I_L(x, y) - I_R(x + i, y + j) - \hat{\mu}_{i, j}(x, y))^2,$$

$$\hat{\mu}_{i, j}(x, y) = \frac{1}{W} \sum_{i, j} W_{i, j} (I_L(x, y) - I_R(x + i, y + j)),$$

$$W = \sum_{i, j} W_{i, j},$$

$$i \in [(l-1)/2, (l+1)/2],$$

²Among other popular corner detectors is the Moravec's formulation [8].

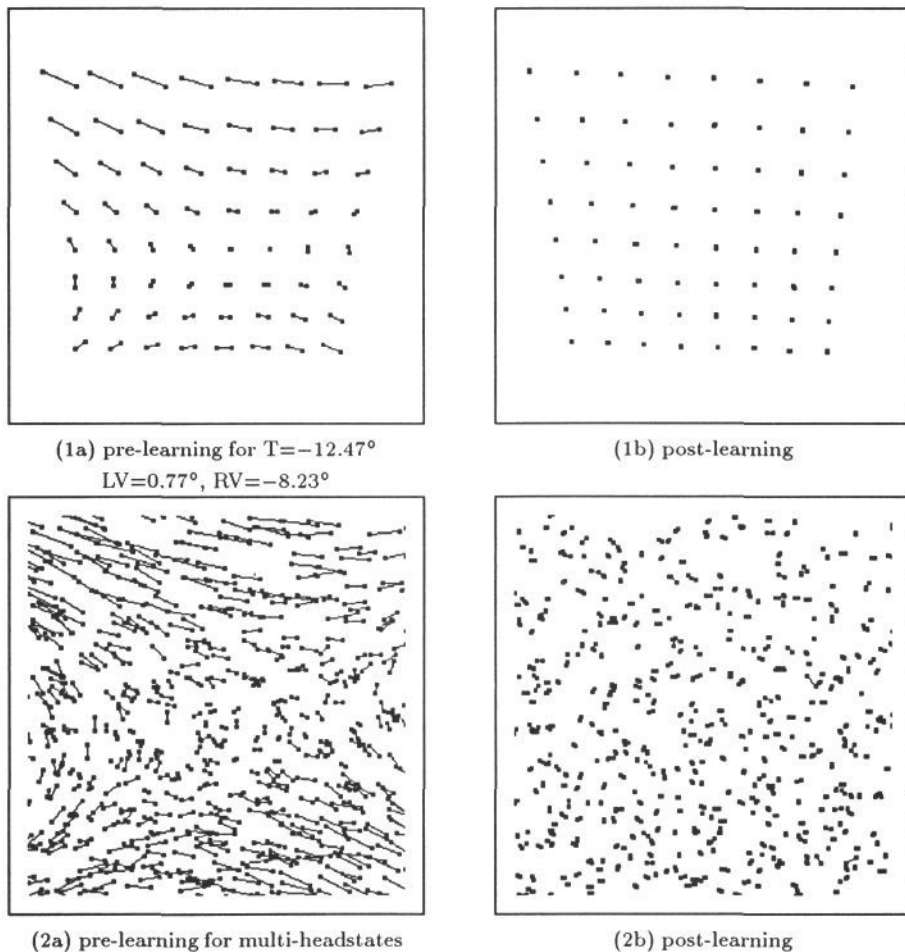


Figure 2: Comparison of pre-learning disparity and post-learning disparity errors

$$j \in [(h-1)/2, (h+1)/2],$$

l: the patch length,

h: the patch height (width),

$W_{i,j} = \left(1 - \frac{|i|}{2l}\right) \left(1 - \frac{|j|}{2h}\right)$: patch masking coefficients which put a priority over pixels near center.

It is easy to see that computed disparity satisfies the maximum similarity between two patches in the left image and the right one respectively.

6 Combination of Prediction and Matching Errors

Now for a corner point $p = (x, y)$ in the left image, we can compute its matching disparity (d_x, d_y) from the correlation. Also we can obtain its prediction disparity

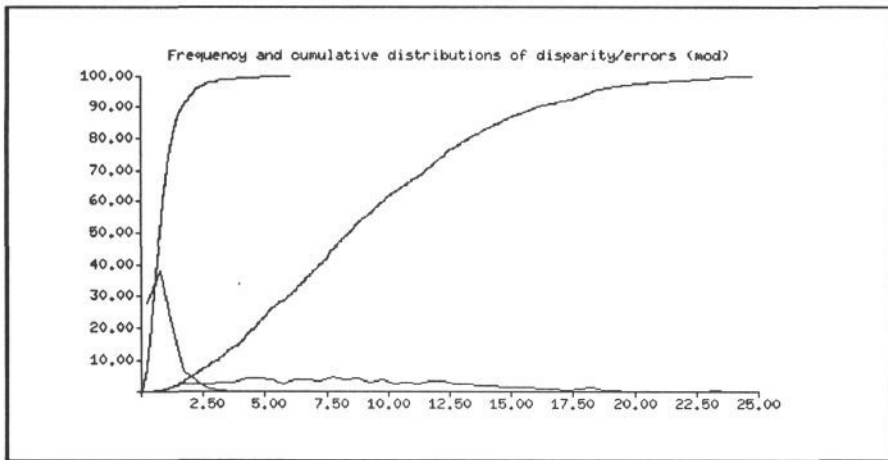


Figure 3: Frequency and cumulative distributions for pre-learning disparity and post-learning errors

(d'_x, d'_y) using PILUT. The difference between those two disparities then serves as an indicator of that the point p comes from an obstacle or the ground plane. However, experiments show that this deviation is not always very robust, especially when we are trying to detect small obstacles. So we need combine it with other measurements. The values for $\hat{\mu}_{d_x, d_y}$ and $\hat{\sigma}_{d_x, d_y}^2$, attained whilst computing the matched correspondence, can be thought as the mean and variance of the gaussian distribution of residual error. Also the mean $\hat{\mu}_{d'_x, d'_y}$ and variance $\hat{\sigma}_{d'_x, d'_y}^2$ for the disparity (d'_x, d'_y) predicted by PILUT at pixel (x, y) can be computed. The similarity between the two distributions $N(\hat{\mu}_{d_x, d_y}, \hat{\sigma}_{d_x, d_y}^2)$ and $N(\hat{\mu}_{d'_x, d'_y}, \hat{\sigma}_{d'_x, d'_y}^2)$, like the disparity deviation, gives as a measure of the probability that the pixel (x, y) represents a point that lies on the ground plane. Here we use the Bhattacharyya distance between two univariate distributions $N(\mu_1, \sigma_1)$ and $N(\mu_2, \sigma_2)$ [1],

$$B = \frac{(\mu_1 - \mu_2)^2}{4(\sigma_1^2 + \sigma_2^2)} + \frac{1}{2} \log \left(\frac{\sigma_1^2 + \sigma_2^2}{2\sigma_1\sigma_2} \right),$$

With the hypothesis H , a given point lies on the ground plane, and bayesian variables D for the disparity deviation and B for the Bhattacharyya distance, it is easy to build a simple casual Bayesian net, illustrated in Figure 4.

Suppose we do not have any knowledge about the prior probability $P(H)$, then we take $P(H) = P(\bar{H}) = 0.5$. Then the probability updating can be described as follows:

$$P(H | D) = \frac{P(D | H)}{P(D | H) + P(D | \bar{H})},$$

$$P(H) \leftarrow P(H | D),$$

$$P(H | B) = \frac{P(B | H) \cdot P(H)}{P(B | H) \cdot P(H) + P(B | \bar{H}) \cdot P(\bar{H})}.$$

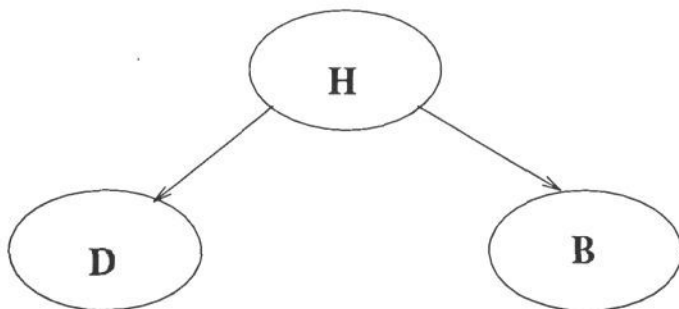


Figure 4: The casual bayesian net for GPOD

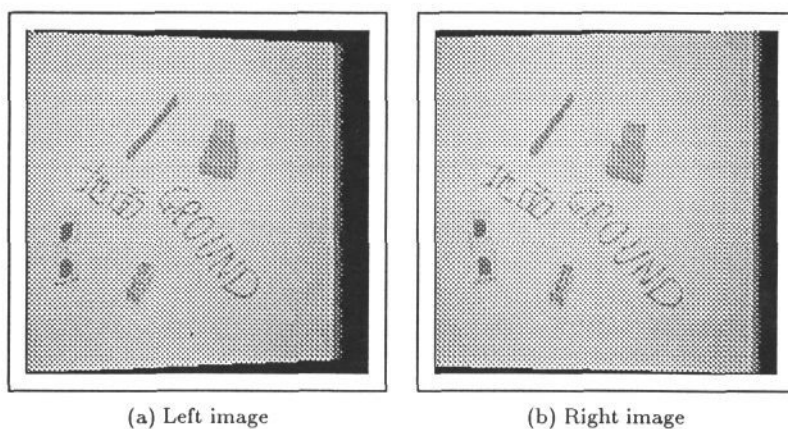


Figure 5: Rectified image pair of floor and obstacles

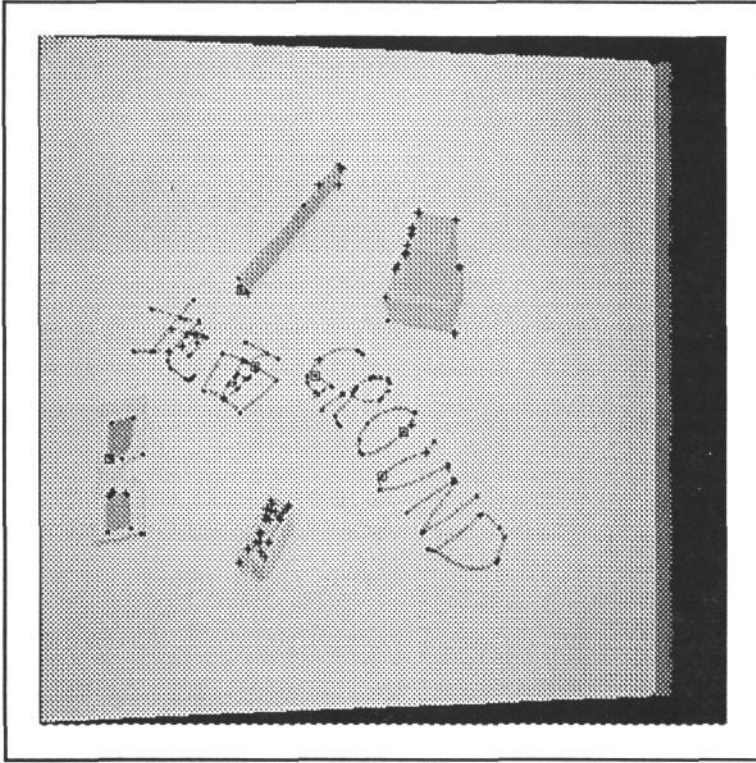
7 Experiment results and Conclusion

The rectified image pairs shown in Figure 5 was obtained to test the above algorithm. The obstacles consist of a piece of card of thickness of about 1mm, a pen with diameter of 7mm, a battery dimensions of $45(l) \times 26(w) \times 15(h)$ mm, and a typical industrial object of bigger size. Also some English letters and Chinese characters were written on the floor in order to make it well textured.

Figure 6 shows the retinal error distributions for ground plane with and without obstacles. Figure 7 illustrates how corners were recognised arising from floor or an obstacle. Figure 8 gives the receiver operating characteristic curve.

From the experiment results, the following conclusions can be drawn:

- *A PILUT using affine fitting blended with Gaussian RBF gives a good fit the disparity flow field mapping across the whole head state dimension.*
- *The probabilistic measurement derived from disparity deviation and Bhattacharyya distance performs well.*



• : a point on ground floor □ : something unsure × : obstacles

Figure 7: Ground plane obstacle detected

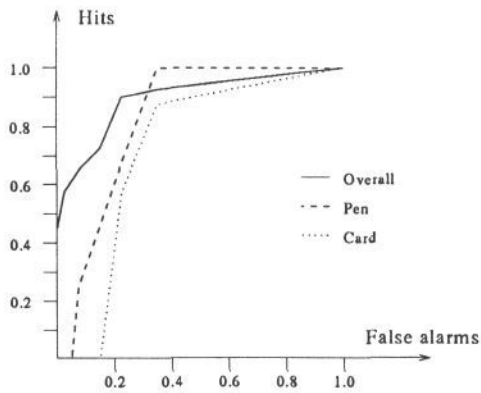


Figure 8: Receiver operating characteristic curves

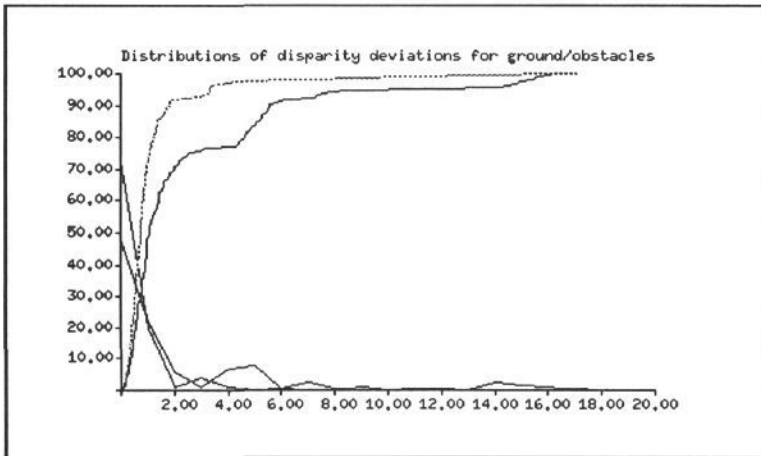


Figure 6: Frequency and cumulative distributions of disparity deviations for ground and ground with obstacles respectively

- *The system can precisely detect the battery and the industrial object, while it keeps the false alarm error less than 5%.*
- *More than 50% corners coming from an as thin as 1mm piece of card can be detected, but at the cost of about 20% false alarm error.*

8 Acknowledgements

Y. Shao is sponsored by the Sino-British Friendship Scholarship Scheme. Authors are indebted to all the members in AIVRU.

References

- [1] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions", *Bull. Calcutta Math. Soc.*, **35**, pp.99
- [2] D. M. Booth, Y. Zheng, J. E. W. Mayhew, M. K. Pidcock, "Ground plane obstacle detection using stereo disparity", AIVRU internal document
- [3] S. M. Cornell, J. Porrill and J. E. W. Mayhew, "Ground plane obstacle detection under a variable camera geometry using a predictive stereo matcher", *Proc. BMVC'92, Leeds, Sept. 1992*, pp.548
- [4] W. Förstner, E. Gülch, "A fast operator for detection and precise location of distinct points, corners, and centres of circle features", *Proc. of Intercom. Conf. on Fast Proc. of Photogram. Data, Interlaken, Switzerland, 1987*, pp.281

- [5] M. R. M. Jenkin and A. Jepson, "Detecting floor anomalies", Proc. 5th BMVC, University of York, 13-16 Sept. 1994, **2**, pp.731
- [6] J. Porrill and J. E. W. Mayhew, "Approximation by linear combinations of basis functions", Feb 6, 1995, AIVRU internal document
- [7] J. E. W. Mayhew, Y. Zheng, S. Cornell, "The adaptive control of a four-degrees-of-freedom stereo camera head", Phil. Trans. R. Soc. Lond. B, **337**, 1992, pp. 315
- [8] H. P. Moravec, "Obstacle avoidance and navigation in the real world by a seeing robot rover", Memo AIM-340, Stanford University, Sept. 1980
- [9] S.D.Hippisley-Cox, S. Harrison, J. E. W. Mayhew, "Stereo under variable camera geometry I", AIVRU internal document, July 1995
- [10] N. A. Thacker, J. E. W. Mayhew, "Designing a layered network for context sensitive pattern classification", Neural Networks, **3**, 1990, pp.291