

Real Time Obstacle Detection for AGV Navigation Using Multi-baseline Stereo

Han Wang

School of Electrical & Electronic Engineering,
Nanyang Technological University, Singapore 639798
e-mail: hw@ntu.edu.sg

Jian Xu Javier I Guzman

Gintic Institute of Manufacturing Technology, Singapore 638075

Ray A Jarvis

Department of Electrical & Computer Systems Engineering
Monash University, Melbourne, VIC 3168, Australia

Terence Goh Chun Wah Chan

Defence Science Technology Agency, 1 Depot Road, Singapore 109679

Abstract: This paper reports the real time vision algorithm ELS, designed for obstacle detection and ground identification for AGV navigation using multi-baseline stereo. Our major contributions are (1) the Disparity Gradient Filter for noise reduction, (2) ground orientation detection using extended Hough Transform, and (3) the establishment of error model for obstacle detection. Experiments show that the algorithm works effectively in unstructured terrain.

1. Introduction

Vision based navigation for mobile robots has been implemented successfully in many places using different approaches [9, 1, 11]. Some are based on corners using structure from motion algorithms such as Droid by Harris [2], some are based on lines by Zhang [13], and some are based on neural network [4]. The CMU Navlab project has been a long success [3, 12] in unmanned guided vehicle navigation. Most of these systems required dedicated hardware such as the Systolic Array Processor (Warp) and DSP chips which are very costly.

Comparing with other vision approaches, multi-baseline stereo [5] has the advantage of offering dense range image at very high accuracy. For example, the *Virtuoso* by Webb *et al.* [6] and the *Digiclops* by PointGrey. They are commercially available using only a Pentium processor. *Digiclops* has been used in indoor mobile robot navigation using clustering techniques [7].

Our project is primarily designed for outdoor AGV navigation and the rich texture environment provides excellent features for the stereo matching algorithm. ELS takes disparity map as its input which is computed using the

Sum of Squared-Difference (SSD) algorithm derived by Okutomi and Kanade [8]. Three cameras are used to form two baselines. Stereo images are taken simultaneously and no attempt is made for temporal integration. This gives us the advantage of detecting moving obstacles such as tree leaves and human figures.

The paper is organised following the sequence of the ELS algorithm. Firstly, the noise reduction method of disparity gradient limit is described. After the interpolation, the extended Hough Transform is introduced to find the ground orientation (slope). Finally, we give our conclusions, discuss existing problems and propose future works.

2. Filtering Using Disparity Gradient Limit

It is observed that large amount of noise exist and they are clustered due to the minimisation technique of SSD. Many low pass filters have been used such as the 1-D and 2-D median filter with little success. Further action is needed to clear them.

Assuming the world is continuous, a physical point must have its immediate neighbours or adjacent points. In another words, for any two neighbouring points, their difference in disparity must be sufficiently small. The disparity gradient is defined as

$$DG = \sum_{i \in w} \frac{|D - D_i|}{d_i}, \quad (1)$$

where w is a local window, D is the disparity of the pixel in concern, D_i is the neighbour's disparity and d_i is the *Euclidean* distance. A point is rejected if its disparity gradient is above a certain threshold (or DGL). DGL was first found in PMF [10] by Pollard *et al.* for disparity computing in a relaxation manner. We extend this idea in noise reduction. Figure 1 and 2 show the range points before and after the Disparity Gradient Limit filtering.

However, this method does not work well for occlusion because DGL is based on continuity of matter. Occlusion error has been largely eliminated by removing regions that have very high contrast.

3. Interpolation

SSD relies on texture. In some area, texture is not available (see Figure 3). Interpolation has to be carried out to fill up the gaps. We divide the region into 20x20cm grids. Inside each grid, the mean is used for interpolation. A blank grid is filled under the condition that it has more than one adjacent valid neighbours and the neighbouring grids are in opposite position. This is to prevent the algorithm from extrapolation. The result is shown in Figure 4.

4. Ground Orientation Detection

Extended Hough Transform is used to detect the ground orientation. The ground orientation is effectively a unit normal vector which is only two dimensional. For each 20x20 cm² grid, four facets produces four normals. Each normal is accounted for in the Hough Transform.

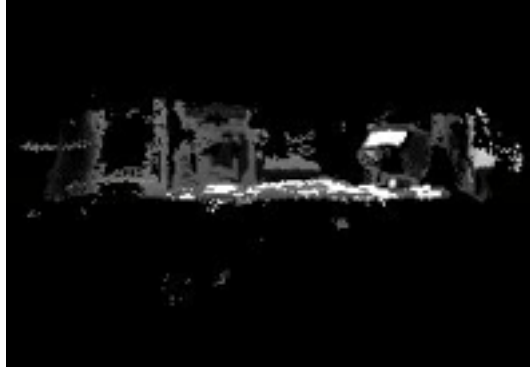


Figure 1. Range image of an office scene. The view point is set at the floor. Noise are visible under the floor (horizontal cluster in the middle of the image).

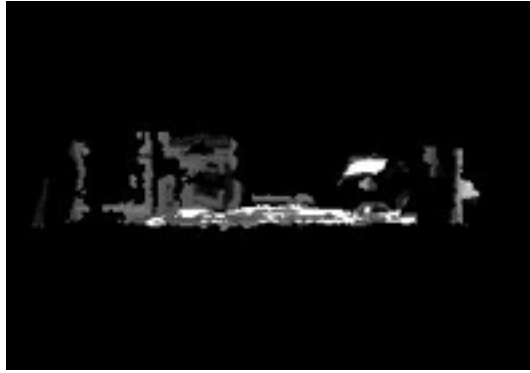


Figure 2. Results after the DGL filter

Surface normal is a unit vector of two degrees of freedom. Let

$$\mathbf{n} = (n_1, n_2, n_3)^T$$

represent the surface normal computed from one facet. It can be rearranged by dividing the vector with n_3 , thus

$$\mathbf{n} = n_3 \left(\frac{n_1}{n_3}, \frac{n_2}{n_3}, 1 \right)^T$$

Setting $\frac{n_1}{n_3} = \alpha$ and $\frac{n_2}{n_3} = \beta$, the parametric space of (α, β) is defined and the extended Hough Transform can be applied.

Figure 5 shows the detected floor orientation with a drawing pin.

5. Error Model and Obstacle Detection

Obstacles are defined as object of 0.4m in height including negative obstacles such as pits and trenches. We define the error model for obstacle detection.

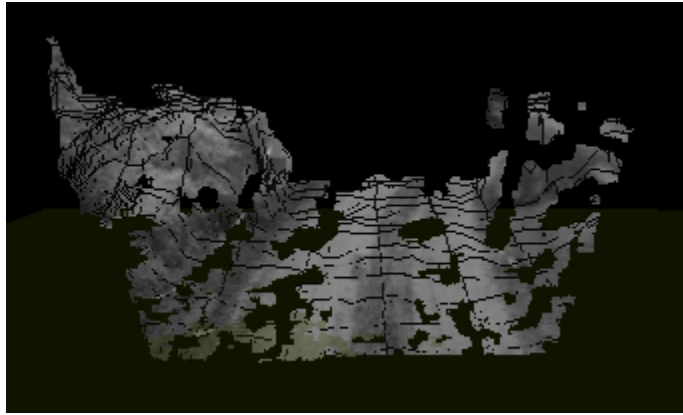


Figure 3. 3D reconstruction of ground floor with texture mapping. Spurious points are filtered out by the disparity gradient filter. In the region where low texture are detected, range information is not available (shown as black holes). The artificial lines are created by the graphics rendering.

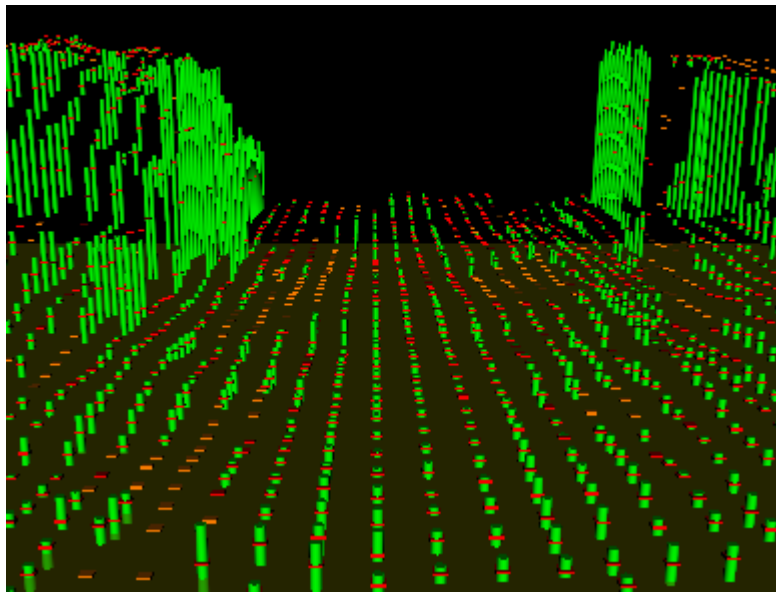


Figure 4. The ground is divided into 20×20 cm² grids. The cylinder shows the maximum and minimum height of the grid with mean marked as a box. The blank regions are filled using linear interpolation techniques.

Given the image resolution w , baseline b and focal length f , for parallel

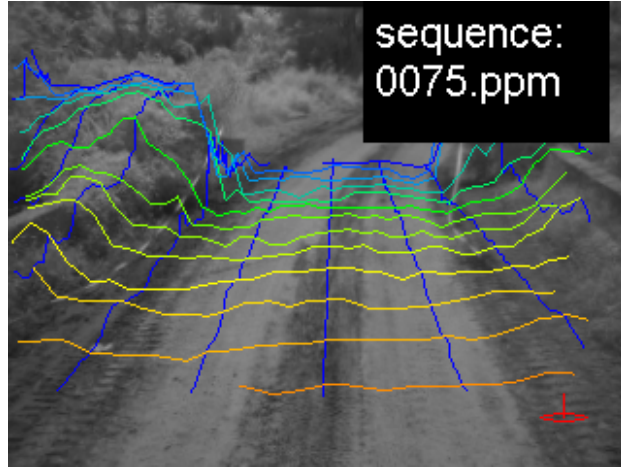


Figure 5. A 3D mesh has been placed on the original grey level image. The mesh shows $1 \times 1 \text{ m}^2$ squares. The drawing pin shows the orientation of the ground.

stereo, the error in depth at distance Z is defined as,

$$e_z = \frac{Z^2}{bf - Z} \quad (2)$$

where f is related with w by the horizontal field of view v (angle)

$$f = \frac{w}{2 \tan \frac{v}{2}}.$$

Assuming square pixels, the error in X (pointing right) and Y direction (pointing down) is given in

$$e_x = \frac{XZ}{fb - Z}, \quad e_y = \frac{YZ}{fb - Z}. \quad (3)$$

Let $\mathbf{e} = (e_x, e_y, e_z)^T$, the probability of an observed 3D point being in error is in Gaussian distribution,

$$P(\mathbf{e}) = e^{-\frac{1}{2} \mathbf{e}^T C^{-1} \mathbf{e}} \quad (4)$$

where C is the observation covariance. \mathbf{e} defines a volume encapsulates an observed 3D point. It is clear that the confidence of measurements decreases linearly as the distance in depth increases.

Table 1 shows the typical error of 0.4m object when the baseline $b = 10 \text{ cm}$ and $w = 320$ with 1/2 sub-pixel accuracy.

Range	Field of View		
	30°	60°	90°
5	.0175	.0397	.0741
10	.0366	.0881	.1818
15	.0575	.1484	.3529
20	.0805	.2258	.6667

Table 1. Error of 0.4m object at different distance.(unit=metres)

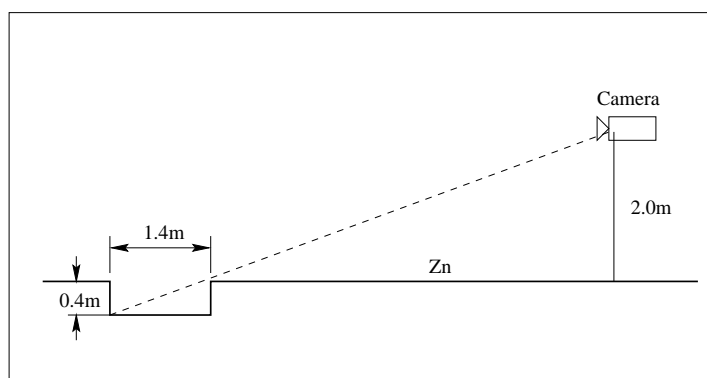


Figure 6. Negative obstacles can be identified at distance Z_n

5.1. Negative Obstacles

We need to identify negative obstacles as well. Figure 6 shows the relation of the distance Z_n from the camera to the pits of 1.4m wide. The camera is mounted at 2.0m above the ground and the identifiable size is 0.4m (obstacle size). Therefore, the distance can be found as

$$Z_n = \frac{2.0 \times 1.4}{0.4} = 7.0m$$

If the camera is mounted at higher location, the distance of detection is longer and it is more desirable, because more time is given to the navigation module. However, higher camera location will result in unstable imaging due to the rough terrain and longer distance means less confidence in data accuracy.

6. Performances

The system comprises of two Pentium-II 550MHz computers, one is dedicated to SSD and the other is for running ELS. It runs at 4Hz for 320x240x8 images and faster speed is expected with a better processor. It is able to detect obstacle of 0.4m at a distance close to 20 metres.

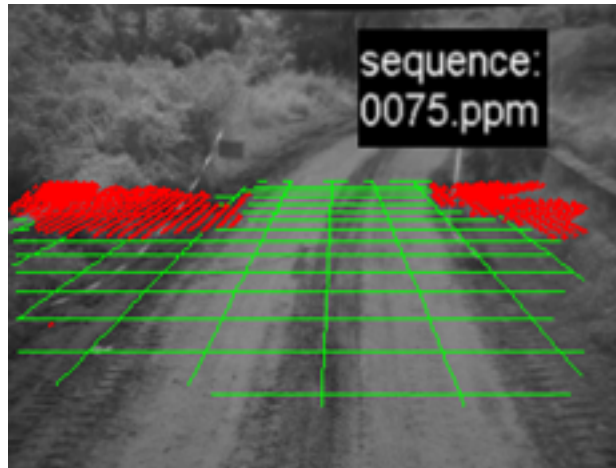


Figure 7. Obstacles are shown in red cross with 20cm grids and flat area is shown in 1m grids.

7. Conclusions and Future Work

DGL is effective in noise reduction and it outperforms traditional low-pass filter in this application. The error model of obstacle enables us to detect obstacle at very early stage. The use of extended Hough Transform makes reasonable estimation of the ground orientation. Further work is in progress to split the road into smaller region for early detection of slopes.

This is an on-going project. A speed of 10Hz is desirable for the robot to travel at 20Km/h. Some optimisation are required to further speedup the process. In addition, two sets of stereo cameras are in consideration in order to cover 90 degrees of horizontal field of view. Presently, the camera is fixed on the vehicle, when the vehicle reached the hill crest, the camera needs larger tilt-down angle. We are actively in search of techniques to solve this problem.

8. Acknowledgements

H Wang would like to thank Department of Electrical and Computer System Engineering, Monash University for providing him with the opportunity and research facilities during his study leave. Discussions with David Suter and Lindsay Kleman have been very beneficial.

References

- [1] S. Bohrer, T. Zielke, and V. Freiburg. An integrated obstacle detection framework for intelligent cruise control on motorway. In *Proceedings of Symposium of Intelligent Vehicles*, 1995.
- [2] C. Harris. Determination of ego-motion from matched points. In *Proc. 3rd Alvey Vision Conference*, pages 233–236, 1987.
- [3] M. Hebert and T. Kanade. 3-D vision for outdoor navigation. In *Image Under-*

- standing Workshop*, pages 593–601, Cambridge, MA, April 6-8, 1988. Morgan Kaufmann.
- [4] T M Jochem, D A Pomerleau, and C E Thorpe. Vision guided lane transition. In *IEEE Symposium on Intelligent Vehicles*, Detroit, September 1995.
 - [5] T. Kanade and T. Nakahara. Experimental results of multibaseline stereo. In *IEEE Special Workshop Passive Ranging*, Princeton, NJ, Oct. 1991.
 - [6] S B Kang, J A Webb, C L Zitnick, and T Kanade. A multibaseline stereo system with active illumination and real-time image acquisition. In *International Conference on Computer Vision*, Cambridge, MA, 1995.
 - [7] D Murray and J Little. Using real-time stereo vision for mobile robot navigation. *Autonomous Robots*, pages 161–171, 2000.
 - [8] M. Okutomi and T. Kanade. A multi-baseline stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(4):353–363, 1993.
 - [9] L. Robert, M. Buffa, and M. Hebert. Weakly-calibrated stereo perception for rover navigation. In *Proceedings of International Conference on Computer Vision (ICCV95)*, 1995.
 - [10] Pollard SB, Mayhew JEW, and Frisby JP. PMF: A stereo correspondence algorithm using a disparity gradient limit. *Perception*, 14:449–470, 1985.
 - [11] S. Singh and B. Digney. Autonomous cross-country navigation using stereo vision. Technical Report CMU-RI-TR-99-03, Carnegie Mellon University, 1999.
 - [12] C. Thorpe, S. Shafer, and T. Kanade. Vision and navigation for the Carnegie Mellon Navlab. In *Image Understanding Workshop*, pages 143–152, Los Angeles, CA, Feb. 23-25, 1987. Morgan Kaufmann.
 - [13] Z. Zhang. Estimating motion and structure from correspondence of line segments between two perspective images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(2):1129–1139, 1995.