Transactions on Information and Communications Technologies vol 16, © 1996 WIT Press, www.witpress.com, ISSN 1743-3517

Obstacle detection using stereo without correspondence

L. X. Zhou & W. K. Gu
Institute of Information & Intelligent System, Department of Information & Electronic Engineering, Zhejiang University, Hangzhou 310027, P. R. China

Abstract

Stereo vision is a straightforward approach for 3D information perception. However image correspondence is the most difficulty in real-time applications. In this paper a qualitative 3D scene verification method is proposed. The global correspondence problem is simplified to an optimization for disparity map parameters. The results of fitting provide the camera pose parameters. The obstacles are then detected from the abnormality of disparity. As we known, simple two-dimensional vision based road following methods are not sufficient for Autonomous Land Vehicle (ALV) navigation in complex environments under arbitrary weather and illumination conditions. They can not recognize 3D obstacles efficiently, and may detect the shadow or water on the road as false obstacles. In this paper the novel qualitative stereo-vision method we proposed is designed for real-time obstacle detection on structural roads. The optimal disparity map of the image pair can be easily computed, and it is a linear function on the image plane. Road scene is verified by the optimal disparity map. A morphology procedure is applied for more reliable abnormal-disparity region extraction. These regions are the focus-of-attention parts for following processing, as obstacle avoidance or object recognition. The algorithm is not correspondence related, so it is very efficient and can be implemented in real-time. Experiments show that this approach is simple and robust.

Transactions on Information and Communications Technologies vol 16, © 1996 WIT Press, www.witpress.com, ISSN 1743-3517

1 Introduction

Autonomous Land Vehicle (ALV) road following and obstacle avoidance require real-time algorithms. These tasks usually rely on special-purpose hardware. Road following methods can be divided into two categories, which are region-based and edge-based [1-4]. In our test environment, there are relatively clear boundaries between the road and non-road regions. We have developed an efficient real-time lane mark line following system for ALV navigation [9]. The following procedure can be implemented on a TMS320C30 based image processing board or a personal computer.

The purpose of ALV obstacle detection is to find concave or convex objects on the ground plane. The significant feature of these objects is that they do not satisfy the smooth constraint of the ground. Previous ALV systems used to fuse the 2D vision system and the 3D active ranging methods. For instance, our ALV utilized the 3D laser radar to obtain the complete depth maps. However the active equipment is quite expensive, and a cheaper, reliable substitution must be found.

There have been some obstacle detection approaches, as optical-flow based or image re-projection based. These schemes all assume that the vehicle is running on a smooth and straight road without bumping. These algorithms require the vehicle situation parameters and they are not real-time as well.

Binocular stereo vision is the most straightforward approach in machine depth perception. Although there are some strong constraints, as epipolar constraints, in stereo vision, the correspondence problem is always regarded as a global optimization procedure [8]. Therefore the stereo matching is always the difficulty for its real-time applications. For example, Reference [5] presents the NASA planetary mobile robot using correlation-based stereo. It can only be implemented near real-time and depends on special correlation hardware. Another obstacle detection method using stereo image pair without matching is present in [6]. However it is indeed reprojection based and requires fixed calibration parameters beforehand. It also assumes that the vehicle egomotion is smooth and steady.

In fact, ALV obstacle avoidance is not a general-purpose stereo task. It does not require a complete depth map like that of a laser radar. It only requires the qualitative information about whether there is or is not an object. Because the vehicle is perhaps bumping and rolling while going, the disparity map parameters are varying at all time. In this paper the parametric model of the smooth road is established based on a binocular stereo system. The global correspondence problem is then changed to a model-based parameter optimization of the disparity map. It is the inverse problem of stereo matching. Real road scene is verified by the optimal disparity map. Then the abnormal regions of the image are extracted.

Transactions on Information and Communications Technologies vol 16, © 1996 WIT Press, www.witpress.com, ISSN 1743-3517

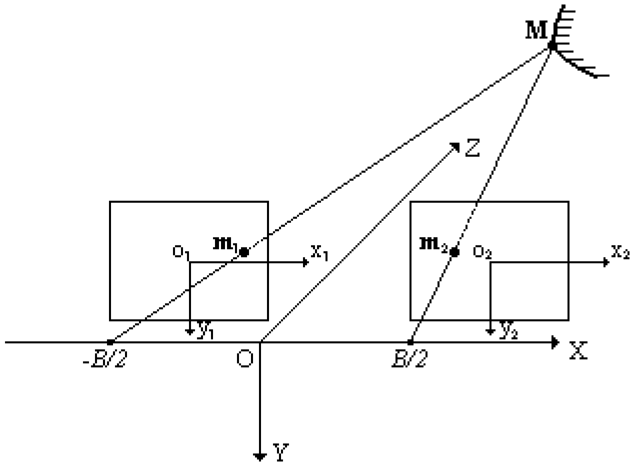

Figure 1: Parallel-setup stereo system coordinates

2 Algorithm description

The parallel axis stereo system coordinates are shown in Figure 1. Assume the projections of a space point $\mathbf{M}[X,Y,Z]^T$ are $\mathbf{m}_1[x_1,y]^T$ and $\mathbf{m}_2[x_2,y]^T$ respectively. The inverse projection form 2D to 3D is

$$\mathbf{M} = [X,Y,Z]^T = \left[\frac{B(x_1+x_2)}{2(x_1-x_2)}, \frac{By}{x_1-x_2}, \frac{Bf}{x_1-x_2}\right]^T \tag{1}$$

where f is the focus length, B is the baseline, and the disparity $\Delta = x_1 - x_2$.

For the ease of disparity computations, the so-called disparity map coordinates is introduced:

$$\begin{aligned} x &= \frac{x_1+x_2}{2} \\ y &= y_1 = y_2 \end{aligned} \tag{2}$$

Therefore Equation (1) turns to

$$\mathbf{M} = [X,Y,Z]^T = \frac{B}{\Delta}[x,y,f]^T \tag{3}$$

Regard the road as a plane defined as

$$\Pi:\quad Y = \alpha X + \beta Z + H \tag{4}$$

where H is the height of camera, H and α, β are all unknown parameters.

If some reference points on the road have been known,

$$\mathbf{B}_i = [X_i, Y_i, Z_i]^T \qquad i = 1,2,\ldots,N$$

The procedure of finding the optimal fitting plane is to minimize

$$\varepsilon^2 = \frac{1}{2}\sum_{i=1}^{N} d^2(\mathbf{B}_i, \Pi) \tag{5}$$

where $d^2(\cdot)$ is a certain distance measure. In Ref. [7] we have detailed the 3D quantization error properties in a binocular stereo system. It concludes that the localization uncertainties of further depth are far large than those of the small depth. If we choose the Euclidean distance as the measure, i.e.,

$$\varepsilon^2 = \frac{1}{2}\sum_{i=1}^{N} (Y_i - \alpha X_i - \beta Z_i - H)^2$$

then extra errors will be included. Generally speaking, the Mahalanobis distance should be used in above fitting procedure. However its computational cost is quite high. According to the results of [7], we know that the uncertainty of the distance between the fitting point to the fitting plane, increases direct proportionally with Z (the depth of object) or inverse proportionally with Δ (the disparity of the image). Thus we choose the following weighted distance for fitting:

$$\varepsilon^2 = \frac{1}{2}\sum_{i=1}^{N} w_i^2 (Y_i - \alpha X_i - \beta Z_i - H)^2$$

where the weights are disparity-related:

$$w_i^2 = \Delta_i^2$$

Therefore the optimization criterion becomes

$$\varepsilon^2 = \frac{1}{2}\sum_{i=1}^{N} \Delta_i^2 (Y_i - \alpha X_i - \beta Z_i - h)^2 \tag{6}$$

From Equation (3) and (6) we have

$$\varepsilon^2 = \frac{B^2}{2}\sum_{i=1}^{N} (y_i - \alpha x_i - \beta z_i - h)^2 \tag{7}$$

where

$$h = \frac{H}{B}$$

ε^2 reaches its minimum if and only if

$$\begin{aligned}
\frac{\partial \varepsilon^2}{\partial \alpha} &= B^2 \sum_{i=1}^{N} (y_i - \alpha x_i - \beta z_i - h)(-x_i) = 0 \\
\frac{\partial \varepsilon^2}{\partial \beta} &= B^2 \sum_{i=1}^{N} (y_i - \alpha x_i - \beta z_i - h)(-z_i) = 0 \\
\frac{\partial \varepsilon^2}{\partial h} &= B^2 \sum_{i=1}^{N} (y_i - \alpha x_i - \beta z_i - h)(-1) = 0
\end{aligned} \tag{8}$$

Transactions on Information and Communications Technologies vol 16, © 1996 WIT Press, www.witpress.com, ISSN 1743-3517

The unknown parameters α, β, h are immediately solved from above equations.

We want to find the optimal disparity map, or the disparity distributions of the optimal plane on the image. Consider the following problem, which is to find the intersection point of the plane Π and the eyesight:

$$\begin{cases} Y = \alpha X + \beta Z + H \\ [X, Y, Z]^T = \tau \cdot [x, y, f]^T \end{cases} \tag{9}$$

The unknown parameter τ is solved:

$$\tau = \frac{H}{y - \alpha x - bf}$$

On another hand,

$$\Delta = \frac{Bf}{Z} = \frac{B}{\tau}$$

Therefore the optimal disparity map becomes

$$\Delta = \frac{B}{H}(y - \alpha x - bf)$$

$$\therefore \quad \Delta = \frac{y - \alpha x - bf}{h} \tag{10}$$

It is a linear function on the image plane.

3 Implementation

In order to find the road parameters α, β, H, the reference point $\mathbf{B} = \{\mathbf{B}_i\}$, i = 1, 2,..., N must be determined before hand. It's natural to choose the road boundary points in an ALV system. In a general-purpose stereo vision system, It may be implemented by some complex relaxation correspondence procedures [8]. It's lucky that we have developed a real-time lane mark line following algorithm used for ALV navigation. As shown in Figure 2, road boundaries are traced by a series of square windows with some attributes, as the primary direction *orientation*. According to the epipolar constraints, the pixels in left and right image on the same scanning line and on the same side are candidate matches. If

1) The include angle between Win_L*.orientation*, Win_R*.orientation* and the vertical line are all less than $45°$;
2) The difference between Win_L*.orientation* and Win_R*.orientation* is less than $15°$;

then the boundary positions $\mathbf{m}_L$ and $\mathbf{m}_R$ are localized by a one-dimensional Sobel operator. $\mathbf{m}_L$ and $\mathbf{m}_R$ are immediately matched. Therefore corres-pondence, the most difficult problem in stereo, is avoided.

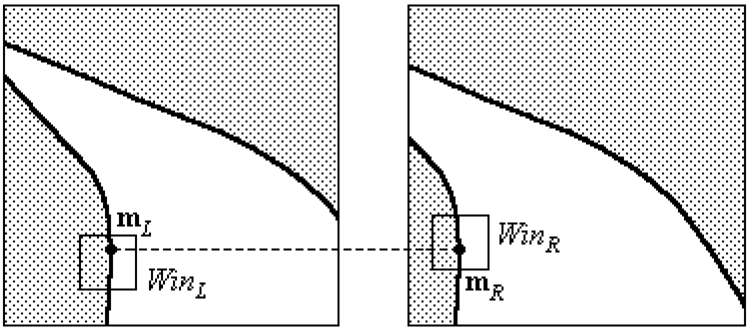

Figure 2: Determine matched edge pixels

The experiment is based on a model. The stereo image of the road is illustrated in Figure 3. The followed boundary is shown in Figure 4. In the experiment we get

$$\alpha = -0.041, \quad \beta = -0.027, \quad h = 6.29$$

The optimal disparity map is obtained using Equation (10), as shown in Figure 5. Check each pixel, G_L and G_R, in left and right image with the disparity that they "should" have. The point pair is defined as an *abnormal pair* if

$$| G_L - G_L | > \theta \tag{11}$$

where θ is a pre-determined threshold. The extracted abnormal regions on the road are shown in Figure 6.

A morphology procedure is applied for more reliably region extraction. Because the obstacles usually have some vertical edges, a small vertical element $[1,1,1]^T$ is used to do an erosion and a dilation operation in succession. Figure 7 shows the result. On the original image a obstacle is detected if there was some black regions of high density, as shown in Figure 8.

Because the most difficult correspondence problem is avoided in our means, the entirely processing time for obstacle detection is only 165 ms on a 486DX2/66 based personal computer.

4 Conclusion

As we known, simple two-dimensional vision based ALV road navigation approaches are not sufficient in complex environments, under arbitrary weather and illumination conditions. For instance, they can not recognize 3D obstacles efficiently; and they may detect the shadow or water on the road as false obstacles. In this paper we present a new qualitative stereo-vision method for real-time obstacle detection on structural roads. The obstacles are labeled by verifying the abnormal-disparity regions on the image. These regions are the focus-of-attention parts for following processing, as obstacle avoidance or object recognition. The algorithm is not correspondence related, so it is very efficient and can be implemented in real-time.

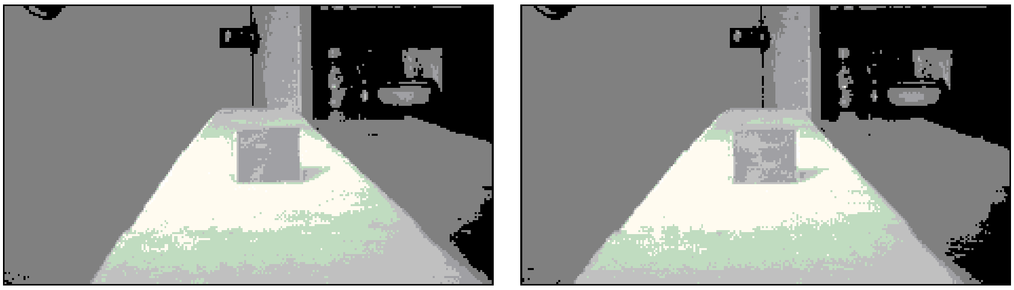

Figure 3: Stereo image of a road model

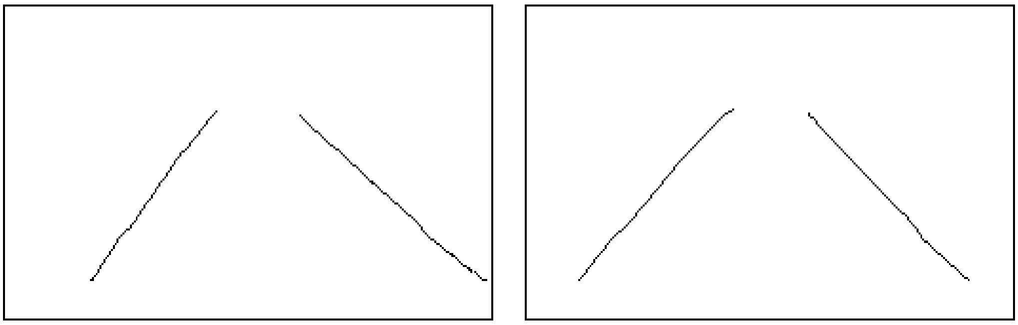

Figure 4: Road boundary obtained by a following program

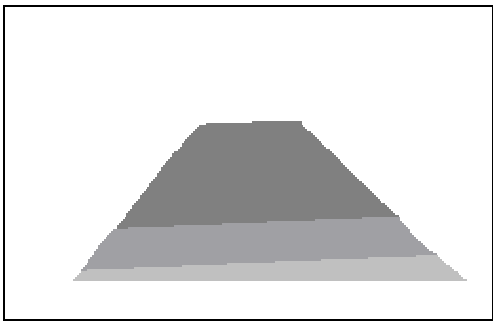

Figure 5: Optimal disparity map

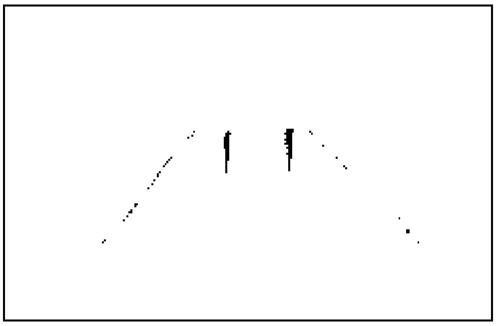

Figure 6: Disparity-abnormal regions

Figure 7: Filtering by morphology

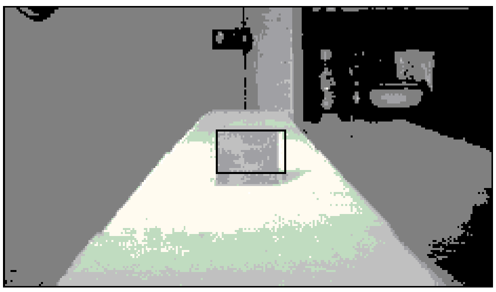

Figure 8: Detected obstacle

References

1. Dickmann, E. & Zapp, A. A Curvature based scheme for improving road vehicle guidance by computer vision, *Proceedings SPIE* Vol. 727, Mobile Robots, 1986.
2. Kuan, D., Phipps, G. & Hsueh, A. C., Autonomous robotic vehicle road following, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **10**, 1988.
3. Thorpe, C., Hebert, M. H., Kanade, T. & Shafer, S. A., Vision and navigation for the Carnegie-Mellon Navlab, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **10**, 362-373, 1988.
4. Gu, W. K., Liu, H. & Wang, H., A robust image processing approach used to ALV road following, *Proceedings 11th International Conference on Pattern Recognition*, Vol. A, 676-679, 1992.
5. Matthies, L., Stereo vision for planetary rover: stochastic modeling to near real-time implementation, *International Journal of Computer Vision*, **8**, 71-91, 1992.
6. Xie, M., Ground plane obstacle detection from stereo pair of image without matching, *Proceedings 2nd Asian Conference on Computer Vision*, Vol. II, Singapore, 1995.
7. Zhou, L. X. & Gu, W. K., Modeling spatial quantization error in stereo, *Proc. 2nd National Symposium on Intelligent Information Technology*, Beijing, 157-160, 1995.
8. Zhou, L. X. & Gu, W. K., Local shape similarity limits for stereo matching, *Proceedings 2nd Asian Conference on Computer Vision*, Vol. II, 671-675, Singapore, 1995.
9. Zhou, L. X., Ye, X. Q. & Gu, W. K., ALV structural road following system, accepted by *3rd International Conference on Signal Processing*, Beijing, 1996.