

Stereo-Vision Based Real time Obstacle Detection for Urban Environments

Qian Yu

State Key Laboratory of Intelligent Technology and Systems, Tsinghua University
China
yuqian97@mails.tsinghua.edu.cn

Helder Araújo

Inst. Syst. Robotics, Dept. of Elect. and Comp. Eng., Univ. of Coimbra
Portugal
helder@isr.uc.pt

Hong Wang

State Key Laboratory of Intelligent Technology and Systems, Tsinghua University
China
wanghong@mail.tsinghua.edu.cn

Abstract

Obstacle detection is an essential capability for the safe guidance of autonomous vehicles, especially in urban environments. Most of the existing stereovision techniques for obstacle detection rely on the planar ground assumption. However in urban environments the real-time requirement and the range of distances that must be considered imply that the planar assumption is too restrictive. In order to enhance the reliability of the obstacle detection task, we do not consider the urban roads as rigid planes, but a quasi-plane, whose normal vector has a predictable constraint. Under this flexible road model, we propose a fast, robust method, which is also easy to implement on DSPs and microprocessors. In this paper, we first model the ideal ground plane in the disparity space and extract the potential obstacle points with the quasi-plane constraint estimated by a Random Sampling Consensus (RANSAC) algorithm offline. Then we introduce a watershed transformation to segment isolated obstacles. To avoid the difficulty of the computation in metric space, the whole process is performed in the disparity image. Various experimental results as well as the real time performance and comparison with results from the planar road model approach are presented, showing the advantages of this method.

1 Introduction

Obstacle detection is an essential capability for the safe guidance of autonomous vehicles, especially in urban environments. Most of existing stereovision techniques for obstacle detection for this type of environments assume a static or dynamic plane model. Several difficulties arise from this assumption. First high speed operation requires a strict real time performance and also a large range of distances. These requirements imply that the use of a strictly planar model of the road is inadequate. Second many roads in urban environments are not completely planar in a small range of distances, often with hills and val-

leys, and even with a slight curvature across the road. Moreover typical urban roads lack texture and reliable landmarks. The existence of texture and/or patterns is essential for the dynamic estimation of the plane parameters. Small errors in the plane parameters severely affect obstacle detection.

To improve the reliability of the obstacle detection, we first introduce a quasi-plane model assumption and propose a fast and robust approach, for obstacle detection without depending on the planar road assumption. This method is able to cope with the common road gradients in urban environments and does not require any rigid restriction on the placement of the stereo head (such as a baseline parallel to the road profile [1]). In order to estimate the parameters of the ground plane offline, a RANSAC plane-fitting algorithm is employed. After the application of the procedure for the extraction of the potential obstacle points, a watershed transform is applied to segment isolated obstacles. To avoid the difficulties of dealing with the metric space, the whole process is performed in the disparity image.

The rest of this paper is organized as follows: Section 2 presents the equation of a plane in the disparity space and introduces the quasi-plane model assumption. Section 3 presents the method to extract the potential obstacle points under the quasi-plane assumption in disparity space. Section 4 describes the obstacle segmentation procedure based on the watershed transformation; Section 5 gives experimental results and describes the performance; finally Section 6 contains the conclusions and the description of future work.

2 Scene assumption and model

2.1 Stereo camera model

For an efficient stereo analysis, it is preferable that the epipolar lines are parallel to the scan lines of the camera. This configuration is obtained with both cameras having the same focal length and with paral-

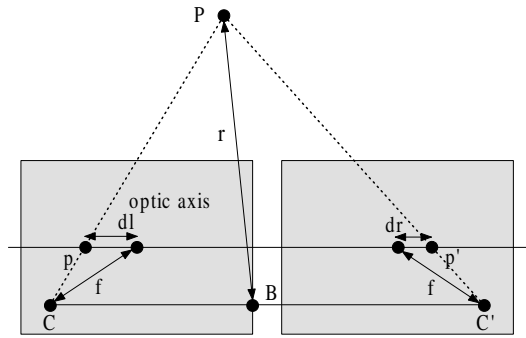


Figure 1: Relationship of disparity and depth

lateral image planes. In these conditions, stereo matching can be performed along the scan lines, as illustrated in Figure 1. The offset of the image location of an object in the left image dl and in the right image dr is called disparity. The disparity is directly related to the distance of the object P from the cameras (measured along the normal to the image planes). With the stereo baseline B and the focus in pixels f , the relationship can be represented as follows:

$$r = \frac{Bf}{d}, \quad \text{where } d = dl - dr \quad (1)$$

In the left camera coordinate system, the image coordinates (using a pinhole camera model) can be expressed as Eq.(2). In a stereo system, we can make the approximation that the pixels are square and therefore the scale factors t_u and t_v are equal: $t = t_u = t_v$. Then if f is the focal length in pixels: $f = f'/t$, where f' is focal length in metric units.

$$\begin{aligned} u &= f \frac{X}{Z} \\ v &= f \frac{Y}{Z} \end{aligned} \quad (2)$$

2.2 A plane in disparity space

In the left camera coordinates, the equation of a plane can be expressed as in Eq.(3), since in our application we do not have to deal with planes passing through the origin, (the left optical center).

$$aX + bY + cZ = 1 \quad (3)$$

On the basis of Eq.(1), (2) and (3), we obtain the equation of a plane in disparity space (d, u, v) as follows:

$$d = B(au + bv) + Bfc \quad (4)$$

From Eq.(4), as we can see, a plane in 3D space is also a plane in disparity space. Note that the partial derivatives of d with respect to u and v are a function

of the first two components (a, b) of the plane normal:

$$\left(\frac{\partial d}{\partial u}, \frac{\partial d}{\partial v} \right) = B(a, b) \quad (5)$$

Using the assumption of a quasi-planar scene it will be shown that the knowledge of both a and b is crucial for the discrimination between the road surface and the obstacles.

2.3 Quasi-planar scene assumption

In the remaining of this paper we will use the two coordinate systems shown in Figure 2: R_w (road) and R_c (left camera). The angle between the optic axis and the road plane is α ; the angle between the X_c axis (which coincides with the baseline) and the X_w axis is β .

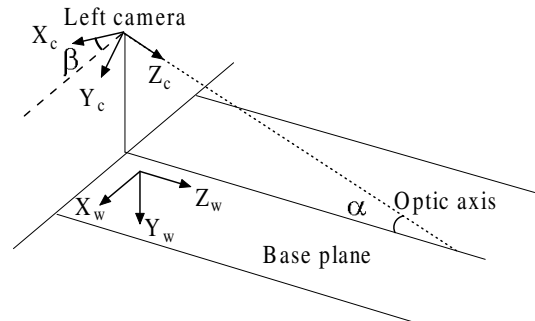


Figure 2: The coordinate systems used

Many roads in urban environments are not completely planar, often with longitudinal (along the road) curvatures, even with slight latitudinal (across the road) curvatures. Instead of a rigid plane, we model the road as a smooth surface, so called a quasi-plane, which is a plane in general and whose normal vector is constrained within a range around the normal of the base plane. Under this assumption, we define obstacles as objects above the road surface, whose normal vector is almost perpendicular to that of local road surface. According to the quasi-planar scene assumption, the unit normal vector of the road surface (a_w, b_w, c_w) can be represented in spherical coordinates as follows (see Figure 3):

$$\{N(R, \theta, \phi) | R = 1, 0 < \theta < \theta_0\} \quad (6)$$

where θ is the latitude angle relative to the negative Y axis. The unit normal vector of the obstacle surface can be represented as follows:

$$\{N(R, \theta, \phi) | R = 1, \frac{\pi}{2} - \theta_1 < \theta < \frac{\pi}{2} + \theta_1\} \quad (7)$$

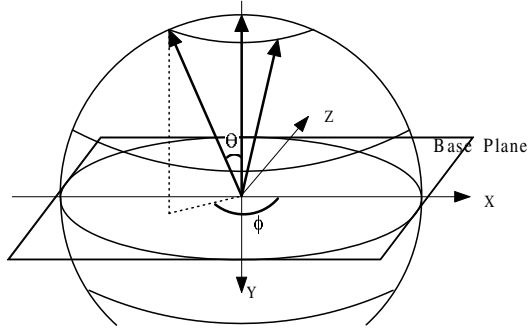


Figure 3: The field of unit normal vector (a_w, b_w, c_w)

3 Obstacle detection

3.1 Normal vector based obstacle detection

The obstacle detection task can be regarded as a procedure to remove points from the road surface. The points to be removed are those whose unit normal vectors satisfy the constraint of the road surface normal 6. Instead of using the three coordinates of the normal vector in 3D space (a_w, b_w, c_w) we will use only the two (a_w, b_w) . Their corresponding components in the left camera coordinate system are computed as in Eq.(8). In the left camera coordinate system, the unit normal vector (a_c, b_c, c_c) can be represented as follows:

$$[a_c, b_c, c_c]^T = R(\beta)R(\alpha) [a_w, b_w, c_w]^T \quad (8)$$

$$\text{where, } R(\beta) = \begin{bmatrix} \cos \beta & -\sin \beta & 0 \\ \sin \beta & \cos \beta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$R(\alpha) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha \\ 0 & \sin \alpha & \cos \alpha \end{bmatrix}$$

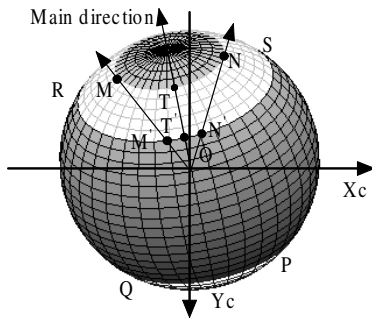


Figure 4: The field of (a_c, b_c) in image plane

The angle θ'_0 , $\angle TON$ (in image plane) and the length of OT (see Figure 4), can be represented respectively as follows:

$$\tan \theta'_0 = ((\cos \alpha / \tan \theta_0)^2 - (\sin \alpha)^2)^{-1/2} \quad (9)$$

$$OT = \cos \alpha \cos \theta_0 - \sin \alpha \sin \theta_0 \quad (10)$$

In Figure 4, the region MTN corresponds to points on the road surface and the region $RSQP$ corresponds to the obstacle surface. If θ_1 in Eq.(7) $< \alpha$, T' is below the X_c axis. The application of conditions Eq.(6) and Eq.(7) is simpler, for that the direction of the vector (a_c, b_c) is sufficient to discriminate between regions of the road surface and the obstacle region. In Eq.(11) a depth ratio is defined. This depth ratio will be used to scale the magnitude of the vector (a_c, b_c) .

$$\text{ratio} = \frac{Z}{dis} \quad (11)$$

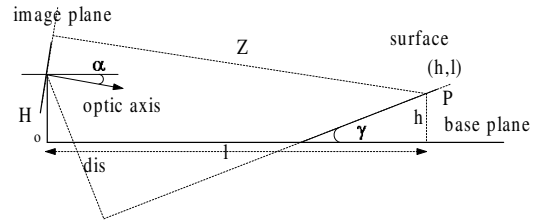


Figure 5: Illustration of the depth ratio

In the sector $OMTN$, the imaging process can be represented approximately as in Figure 5. Z is the depth of point P on the obstacle or road surface; dis is the distance of the optical center to the surface. H is the height of camera relative to the base plane; h the height of point P relative to the base plane and l is the distance between point P and the optical center measured on the base plane. γ is the angle between the surface and the road plane; α is the angle between the optical axis and the base plane. The depth ratio can be expressed in terms of these parameters as follows:

$$\text{ratio} = \left| \frac{(\frac{h-H}{l} - \tan^{-1} \alpha)(1 + \tan^2 \gamma)^{1/2}}{(\frac{H-h}{l} + \tan \gamma)(1 + \tan^{-2} \alpha)^{1/2}} \right| \quad (12)$$

Since $(h-H)$ is usually much smaller than l , $(h-H)/l$ is negligible. Thus we get

$$\text{ratio} = \frac{\cos \alpha}{\sin \gamma} \quad (13)$$

Therefore the depth ratio for regions of the road surface will be larger than the depth ratio for the obstacle surface. For that reason vector (a_c, b_c) is scaled

by the depth ratio defined in Eq.(11). As a result, the difference between the normal vector corresponding to the road surface region $OMTN$ (Figure 4) and the normal vector corresponding to the obstacles surface region $OM'N'$ (Figure 4) is increased. The vector (a_c, b_c, c_c) corresponding to a planar patch defined as in Eq.(3) can be expressed as follows:

$$(a_c, b_c, c_c) = (a, b, c)/(a^2 + b^2 + c^2)^{1/2} \quad (14)$$

And the distance of the planar patch to the origin is

$$dis = 1/(a^2 + b^2 + c^2)^{1/2} \quad (15)$$

From Eq.(1), (11), (14) and (15), we get the disparity space:

$$ratio(a_c, b_c) = \frac{Bf(a, b)}{d} \quad (16)$$

3.2 Obstacle detection in disparity space

From Eq.(5), we can rewrite the scaled normal vector as

$$ratio(a_c, b_c) = \frac{f}{d} \left(\frac{\partial d}{\partial u}, \frac{\partial d}{\partial v} \right) \quad (17)$$

The obstacle detection task can be regarded as a two-step procedure. First, we separate the region $OMTN$ in Figure 4 by using the direction of the scaled vector. Second, the region MTN is discriminated from $OM'N'$ by using the length of the scaled vector. The obstacle detection procedure is shown in Figure 6. We employ an area correlation method implemented by the Small Vision System [2] to acquire dense disparity images, as shown in Figure 6(b). Then we use vertical and horizontal Sobel operators to compute the derivatives of the disparity. Figure 6(e) shows the field of scaled vectors. As we can see in Figure 6(e), the main direction of the base plane is close to 90 degrees. Using an angle (relative to the main direction) smaller than 30 degrees, and a length of the vectors larger than 0.8 (given α is 15 degrees and θ_0 is 30 degrees, computed from Eq.(10) and Eq.(11)), the road surface is removed as shown in Figure 6(c). Finally we use a morphological opening to remove small patches in the detection result (the patches are due to errors in the disparity image). The result is shown in Figure 6(d).

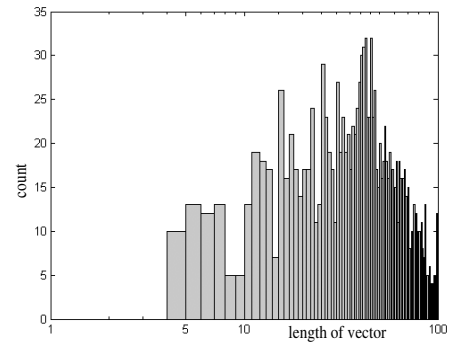
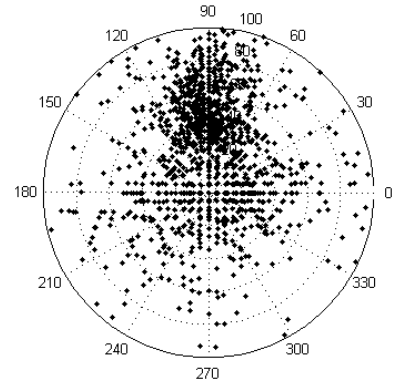
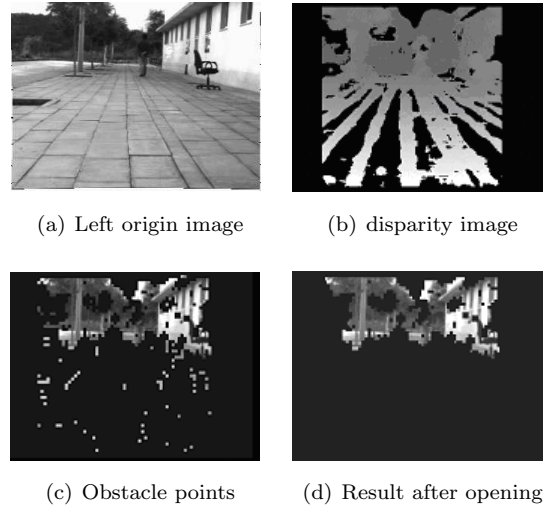


Figure 6: Obstacle detection in disparity space

3.3 RANSAC base plane fitting

In order to obtain the main direction of the base plane of a certain configuration, we propose the use of RANSAC [3], to estimate offline the parameters of the plane defined in Eq.(4). The procedure is the following:

1. In disparity space, randomly select three points to fit a plane. Check, for each and all the points (by computing the distance of a point to the plane), whether they satisfy the estimated plane equation and count the number of fitting points.
2. Repeat step 1 m times, select the triple (a, b, c) with maximum support and do least-squares base plane fitting to this triple using all its supporting points. Provided that there are sufficient road plane points in the disparity space, the plane thus estimated will correspond to the road plane. The probability of complete fitting (without outliers) is given by $1 - (1 - (1 - e)^p)^m$ where e is the contamination fraction, p is the sample size and m the number of samples. In this case the sample size is
3. Assuming that the percentage of contamination for a offline scene is 75%, a probability of 99% for a good sampling requires a number of approximately 300 samplings.



(a) Image of a scene with a grade across the road



(b) Obstacle points with planar assumption



(c) Obstacle points by our method

Figure 7: Comparison with planar assumption: the shade beside the vehicle, which is regarded as obstacle region by mistake in (b), is easily removed in our method.

4 Watershed Obstacle segmentation

As a result of the detection procedure obstacle points are identified. It is then necessary to perform obstacle segmentation. Obstacle segmentation can be regarded as a procedure to cluster similarly connected components in a graph built by the detection procedure. A general image can be regarded as a function $f : D \rightarrow N$, where D is the domain of the image (pixel coordinates) and for some $p \in D$ the value $f(p)$ denotes the value of this pixel, such as gray scale and disparity. In the watershed transformation an image is considered a topographic surface, where $f(p)$ denotes the altitude of the surface at location p . Now, suppose that pinholes are pierced in each local minimum of the topographic surface and the surface is slowly immersed into a lake, shown in Figure 8. Water will fill up the valleys of the surface creating catchment basins. At the pixels where two or more basins would merge we build a "dam". The set of dams obtained at the end of this immersion process is called the watershed transform of the image f .

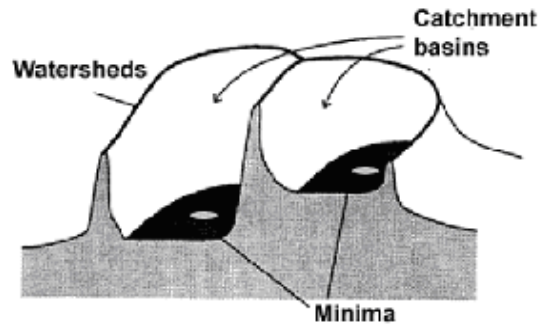


Figure 8: Watershed transformation

If we regard the altitude as similarity of a pixel with its neighbors (connected components), the flooding procedure can be regarded as a clustering procedure that merges the most similar neighbors recursively. In order to segment isolated obstacles, points with similar disparities (which are inversely proportional to depth-see Eq.(1)) are clustered. Thus the obstacles closer to the vehicle will have higher resolution. We can define the similarity of a point to its neighbors as *Laplacian* operator of disparity image:

$$\nabla^2 d(u, v) = \frac{\partial^2 d}{\partial u^2} + \frac{\partial^2 d}{\partial v^2} \quad (18)$$

In practice, we use a non-linear *Laplacian* (NLP) operator [4], which is more robust than Eq.18 and can be implemented by using only integer comparison operations, where $d_n(\cdot)$ is the $n \times n$ square centered



Figure 9: watershed obstacle segmentation

at a point (u, v) :

$$NLP(u, v) = Max[d_n(\cdot)] + Min[d_n(\cdot)] - 2d(\cdot) \quad (19)$$

The computational cost of the watershed transform is related to the number of candidate points and the number of disparity levels (after *NLP* operator). Since the watershed transform is applied only to obstacle regions and since the number of disparities is limited, usually less than 32 (without interpolation), the watershed obstacle segmentation does not affect the real-time requirement. For 320×240 images, the watershed obstacle segmentation takes less than 10ms.

5 Experiments and results

The method proposed in this paper has been implemented in C++ on a commercial Pentium IV 1.4 GHz running under windows 2000. Images of 320×240 resolution are grabbed with an IEEE 1394 interface card by a *MEGAD* digital stereo head. The whole process for detecting obstacle points and segmentation of isolated obstacles is performed in 40ms.

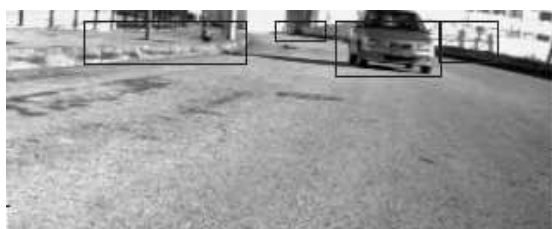


Figure 10: watershed obstacle segmentation

6 Summary and future work

In this paper we first extend the usual planar road model by proposing the flexible quasi-planar road assumption. Then we present a fast and robust method for obstacle detection on flat roads or non-flat roads with slight curvatures. This method uses the sur-

face normal vectors in 3D space. The normal vectors enable the distinction between the obstacles and the road surface. The method is implemented on disparity space. With the results from the detection, obstacle segmentation can be performed by a clustering procedure. This clustering procedure is performed by means of a robust watershed transformation. This approach has several advantages namely:

1. No need for patterns on the road (be planar or non-planar). On the other hand shadows and/or landmarks on the road do not affect the procedure.
2. This method does not impose any specific strict restriction, for example that the baseline of the cameras must be parallel to the road profile.
3. This method is entirely implemented on disparity space. In addition the algorithm can be easily implemented on a standard microprocessor or on a DSP.

This system will be fully integrated on our experimental vehicle used in the Cybercars project. Further improvements will include fusion of intensity information in the obstacle segmentation to achieve a more refined result in visual meaning.

Acknowledgments

This research was supported partly by Chinese High Technology Development Program, by a Portugal-China Science and Technology Cooperation Project and by the European project Cybercars. The author would like to acknowledge, Jianye Lu and Ming Yang (Tsinghua University) for their valuable discussions and comments.

References

- [1] Labayrade, R., Aubert, D., and Tarel, J.-P. "Real Time Obstacle Detection on Non Flat Road Geometry through 'V-Disparity' Representation", "Proceedings of IEEE Intelligent Vehicle Symposium" June 18 2002
- [2] K. Konolige "Small Vision systems: Hardware and Implementation", Eighth International Symposium on Robotics Research, Hayama, Japan, October 1997.
- [3] Stephen Se "Ground Plane Estimation, Error Analysis and Applications Robotics and Autonomous Systems Journal," Volume 39, Issue 2, pages 59-71, May 2002.
- [4] L.J. van Vliet, I.T. Young, and G.L. Beckers "A nonlinear Laplace operator as edge detector in noisy images" Computer Vision Graphics and Image Processing, vol. 45, no. 2, 1989, 167-195.