OBSTACLE DETECTION WITH 3D CAMERA USING U-V-DISPARITY

Yuan Gao, Xiao Ai, John Rarity and Naim Dahnoun

Department of Electrical & Electronic Engineering, Bristol University, Bristol, UK.

ABSTRACT

Obstacle detection has been one of the most critical features for reliable driving scene analysis. This paper presents an approach for an automatic obstacle detection system. The proposed system makes use of depth information generated by a 3D camera mounted on the front of a moving vehicle. Obstacles projected as line features in the V-U-Disparity map can be extracted to detect the road surface and obstacles. A Steerable Filter is employed at an early stage to dramatically lower the noise. Furthermore, a modified Hough Transform is placed to extract the straight line feature from the depth map with improved accuracy. The system is robust in dealing with fault detection caused by roadside features which is a commonly shared problem in many other obstacle detection approaches.

1. INTRODUCTION

Obstacle detection has been an active research field in the automotive industry for decades. The system aims to detect the surrounding on-road obstacles, namely, vehicles and pedestrians, so that collision avoidance, path planning and object classification may be achieved in a driver assistance system. A robust obstacle detection system not only greatly improves the safety but also lowers the false alarm rate and saves the computational cost in the object classification system. Among the various obstacle detection methods, approaches based on stereo-vision have been gaining much research focus. The goal of stereo-vision is to measure the range from the current point of view to an object.

However, in order to extract features such as obstacles and the road surface, further processing needs to be carried out. Many previous works have been done in this field [1][2][3], but they share the same general problems such as uneven road surface, high computational cost by generating reliable disparity map, etc.

Introduced by Labayrade [4], the V-disparity simplifies the extraction of the 3D road surface and obstacles into a 2D linear process. This is achieved by converting the general disparity map into a (u,v) coordinate system.

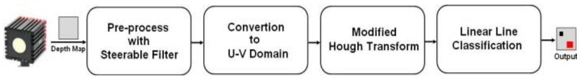

Figure 1. Block diagram of the proposed system.

In the V-disparity domain, longitudinal profile of the road the surface is represented by a piecewise linear curve while the obstacle in the vertical plane is projected as a vertical line. On the other hand, using a similar principle, the U-disparity represented in [5] is capable of projecting the relative side surfaces of the obstacles and structures. In general, the V-disparity would give enough information for separating the road surface and on-road obstacles. However, with the assistance of U-disparity, some false detection such as the road side structures can be eliminated. Furthermore, it may provide richer information of obstacle shape. This will greatly boost the performance of object classification in later stage.

Our proposed system takes depth information generated by a 3D camera. This depth map is pre-processed by a Gaussian based Steerable Filter to emphasise the key disparity features and lower the noise. The filtered disparity map is converted into (u,v) space in the form of V and U disparity spaces. They then undergo a modified Hough Transform to detect possible straight line features which represent obstacles in the real scene. This is followed by classifying linear line candidates according to their position and posture in a U-V-Disparity map. The system block diagram is shown in Figure 1.

2. 3D IMAGING CAMERA

In this study, an alternative to the stereo-vision cameras, we have used the MESA SR4000 3D Imaging camera based on the Continuous-wave (CW) Time-of-flight principle. It is shown in Fig.2 that the active illumination is modulated by a reference sinusoid and emitted by an array of invisible near-infrared LEDs. For each pixel on the focal plane, the reflected light is analysed for its phase delay in order to evaluate the distance; the phase delay is calculated by cross-correlation (mixing and integration) of the reference sinusoid and the received sinusoid. In the actual smart pixel, mixing and integration are done in the charge-domain using CCD technology with low noise. Under sufficient received signal energy, the SR4000 can achieve 1 centimeter distance resolution, compared to the stereo-vision cameras. This enhanced resolution can be

978-1-4577-0690-5/11/$26.00 ©2011 IEEE

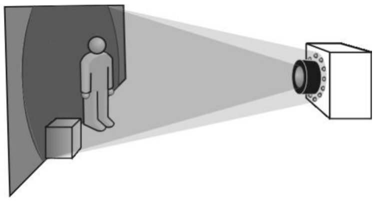

Figure 2. 3D Imaging camera set-up.

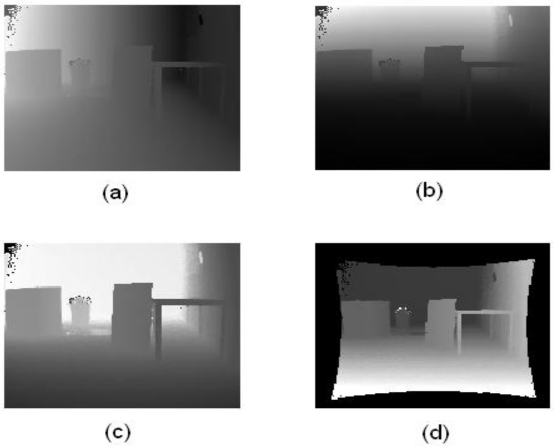

Figure 3. WSC to Disparity conversion: (a) WSC X component; (b) WSC Y component; (c) WSC Z component); (d) Converted disparity map

very beneficial to the U-V disparity based obstacle detection techniques. Also, with the focal plane size of 144x176 pixels, the SR4000 achieves 3D imaging at 50 FPS at up to 10 meters in distance.

3. DISPARITY MAP GENERATION

The Swissranger SR4000 camera [6] generates 3D images in the World Coordinate System (WCS). In order to apply the U-V disparity technique, it is required to be converted into a disparity map; a conversion example is shown in Fig.3. To convert the WCS into the camera perspective, we use the Central Projection [8] as shown in Eq.1.

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} f\,k_x & 0 & x_0 & 0 \\ 0 & f\,k_y & y_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{1}$$

where, u,v is the camera coordinates; x, y, z represent the WCS coordinates; x_0, y_0 denote the coordinates of the principle point; k_x, k_y are the scaling factors; f is the focal length; and

$$x_{CAM} = \frac{u'}{w}; y_{CAM} = \frac{v'}{w} \tag{2}$$

where, x_{CAM}, y_{CAM} are the camera pixel location. The disparity is inverse proportional to the distance as shown in Eq.3

$$d = f\frac{T}{z} \tag{3}$$

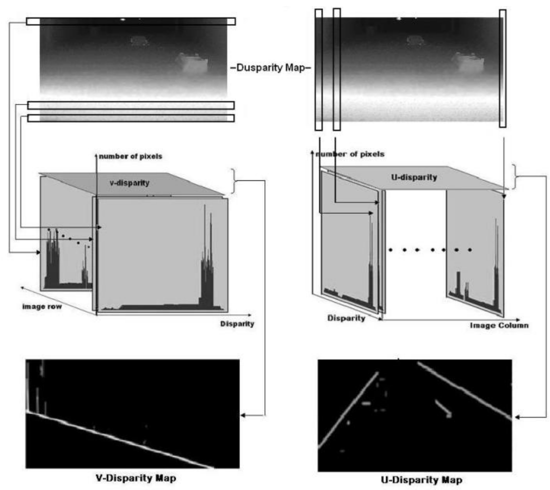

Figure 4. Constructing U-V-Disparity map

where, d is the disparity; f is the focal length; T is the baseline of an assumed stereo-vision camera. The generated disparity map is thus $[x_{CAM} \;\; y_{CAM} \;\; d]^T$.

4. U-V-DISPARITY MAP

In the general stereovision model, the disparity of a point is inversely proportional to the real word distance from the observation point (Eq.3). Derived in [3], the depth map coordinates (*u,v*),the disparity value and the real world coordinate (*X, Y, Z*) follows a simple linear relationship. This gives the principles of the U-V-Disparity map: **(1)** *The plane surface will be projected as a straight line in the V-disparity map***.** **(2)** *Vertical planes in the v-disparity domain are projected as near vertical straight lines.* **(3)** *All side and oblique surfaces parallel to Y-Z plane are projected as straight lines in the U-disparity domain.* By exploiting these properties, the majority of the on-road sense can be projected into the U-V-Disparity domain. Since they all follow a simple linear line distribution, the feature extraction process is greatly simplified.

5. APPLICATION

5.1. Constructing U-V-Disparity Image

U-V-Disparity is built upon a given general disparity map containing the depth value of each pixel. In our proposed approach, the depth map is generated by the 3D camera. To compute the V-disparity map, a histogram of each horizontal row of the disparity map is computed by accumulating the pixels of the same disparity. In each 2-D histogram, the x-axis stands for disparity values from zero to the maximum found within the disparity map, the y-axis is the location of each disparity map row in the corresponding location. The process is illustrated in Figure 4. Following a similar idea, U-disparity is computed by scanning the disparity map by column and accumulating the pixels with same disparity value.

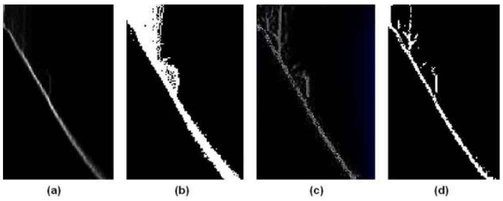

Figure 5. (a) A V-disparity map in practical scene. (b) Available disparity points without any pre-processing. (c) Result of post Gaussian function based Steerable Filter process, with high luminance stands for large likelihood of being on linear lines. (d) Available disparity point post filtering.

Note, the example shown in Figure 4 only demonstrates the basic concept of U and V disparity maps construction. The practical results of these maps are largely distorted, and interfered by noise due to possible disparity map error and complexity of the scene (example shown in Figure 5). In order to extract the straight line features from U and V disparity maps for obstacles and road surface detection, a Hough Transform (HT) is applied to those maps. However, HT is well known for its high sensitivity to noise which dramatically compromises the detection accuracy and efficiency. Therefore, pre-processing to the U and V disparity maps is necessary.

5.2. Steerable Filter

Since we are only interested in linear straight line features formed in the U-V-Disparity map, a Steerable Filter with a base of symmetry Gaussian function is employed. Introduced in [6], the steerable filter applies its base function at many orientations with a small angle rotation each time on an image. Here, the second derivatives of Gaussian are chosen as the basis functions for the steerable filter (shown in Eq.4). These basic functions fit the dark-bright-dark pattern of the lines. The original U and V disparity maps are filtered three times using each of the basic functions below:

$$G_{xx} = \frac{d^2 f(x,y)}{dx^2}; G_{xy} = \frac{d^2 f(x,y)}{dx\, dy}; G_{xy} = \frac{d^2 f(x,y)}{dx\, dy} \quad (4)$$

where f(x) is the Gaussian function. The filter orientation which produces the highest and lowest responses can be calculated using differentiation based on the three filtering results. If a point is on a line, the difference between the highest and lowest filter response will be large. If a point is correspond to noise, the difference between the highest and lowest filter response will not be as significant. Therefore, the difference between the highest and lowest filter response defines the magnitude of directivity. Choosing a threshold that clearly separates the line structure from the rest of the map is then much easier.

The steerable filter gives the map showing confidence of pixels being on a potential continuing line. By using the filtered result, noise pixels in the original map are removed or have their weighting greatly lowered. Figure 5 shows difference between the pre and post results. In our proposed system, to simplify the process, a base Gaussian function is only applied in three directions and the responses in between are interpolated.

5.3. Modified Hough Transform

The Hough Transform is employed in our proposed method to extract straight line features from the U-V-Disparity map. During the experimental stage, we realised that the U and V disparity maps have several characteristics in most of the scenes. In order to optimise the efficiency and accuracy, we exploit these properties in our modified HT. First of all, far field contains more features therefore result in more projection in V disparity map. We set a non-uniform threshold model to guarantee a small cross section tolerance in Hough domain for more accurate detection. Furthermore, far field obstacle tends to lose side and slop detail compared to vertical height. The proposed Hough transform is designed to more focus on searching straight vertical lines in far field and gradually broaden the range while approaching near field with large disparity values. Last but not least, the ground surface plane in most of the road scene shows dominant weighting in V disparity map compared with other obstacles. The straight line detection can then be applied in a course to fine manner: rough scanning to find road surface and then search for possible obstacles with road surface disparity point blinded. In fact, it has been proved in practical testing, that these modifications are almost demanded to maintain a false detection rate at a tolerable level.

5.4. Feature Classification

Once straight line features are extracted from the V-U-Disparity map, candidates are classified according to their properties in table 2 and labeled in Figure 6b and 6c. As a result, lines 1, 2 and 3 are classified as vertical obstacles whilst 4 is the infinity plane. Line 5 in the V-disparity map is projected by the road surface. Line 5 and 6 in the U-disparity represent roadside features. By reverse processing the U-V-Disparity domain features back to original disparity map, pixels in the original map can then be segmented according to their position and depth values (Figure 6d and 6e). These results show that the V-disparity map alone will mistake many road side features as obstacles. Since those side surfaces are detected as environmental side feature in U-disparity, the final result only shows detected obstacle 1, 2 and 3 which are represented in both V and U disparity map disparity map (Figure 6f).

Result of another example scene is shown in Figure 7. The Disparity map (7a) is converted to a V-disparity (7b) and a U-disparity (7c) domain. The final combined result in (7d) indicates that there are obstacles. Fault detection

3 is a vertical surface due to the sunken area caused by a door on the side wall.

Feature Property V/U-Disparity	Classification	Reason
Majority of line contributing points are also the most left/top available disparity point on each row /column.	(1)Ground surface	No point has less depth than ground surface on a given image row/column
Vertical	(2)Infinity	Obstacle with the furthest depth is generally the infinity
Other straight line candidates	(3)Obstacle	-

Table 1. Feature Classification

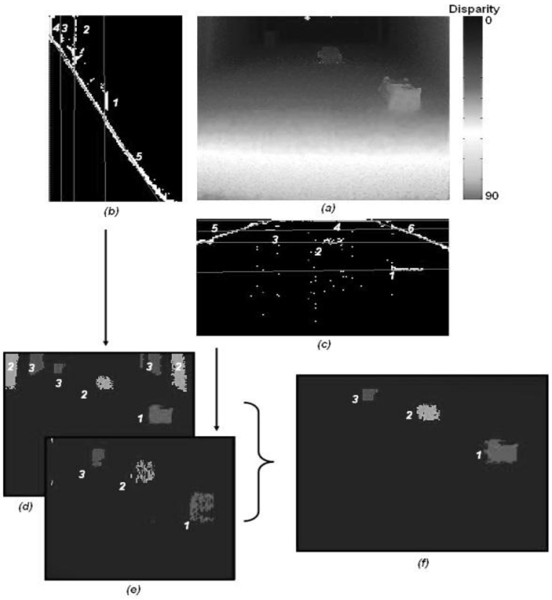

Figure 6. Basic process of obstacles and road surface detection through U-V Disparity.

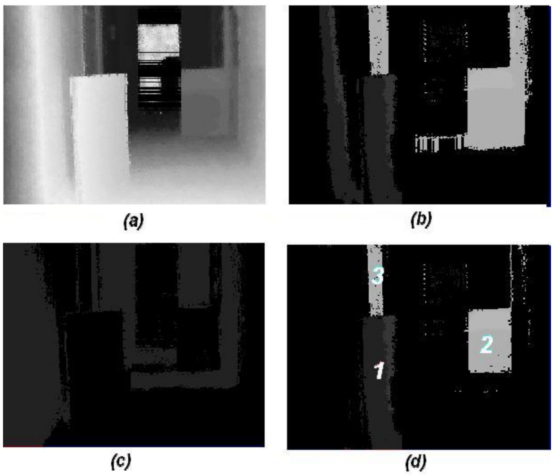

Figure 7. Test example. (a) Disparity map, (b) Result of V-disparity, (c) Result of U-disparity, (d) Final result

6. CONCLUSION

In this paper, an accurate and robust obstacle detection system dedicated for driving scene analysis has been proposed. By employing the depth generated by a 3D camera, obstacles are distinguished from the ground surface by straight line extraction done in a U-V-Disparity map. Without a flat-road assumption, our proposed system is able to work in a more general driving situation. The system also has good tolerance to camera shaking in practical operation. Since the features are extracted by detecting a linear line across the disparity map globally, the system is less sensitive to partial occlusions. The depth map generated through the 3D camera contains much fewer errors than the map produced by traditional stereo-cameras. With the help of a Steerable Filter, irrelevant information is greatly suppressed in order to achieve a faster processing speed and abetter accuracy.

Due to the limitation on hardware installation and lack of evaluation material, the examples and testing process were only carried out under a pre set-up indoors scene. Although the proposed system is believed to perform well under a real drive environment, further testing is still necessary. Furthermore, the current system only employs the U-disparity to eliminate road side features. It is possible to utilise side surface information to outline the obstacle in a more precise shape rather than only from the rear. With obstacle shape information, classification of pedestrian and vehicles may be achieved.

REFERENCES

[1] F. Dornaika, A.D. Sappa, “Real Time Stereo Image Registration for Planar Structure and 3D Sensor Pose Estimation.”

[2] M. Soga, T. Kato, Y. Ninomiya and M. Ohta, “Pedestrain Detection with Stereo Vision,” *IEEE Computer Society,* 2005.

[3] D. Gero´ nimo, A.M. Lo´ pez, A.D. Sappa and T. Graf “Survey of pedestrian detection for advanced driver assistance system,” *Pattern Analysis and Machine Intelligence, IEEE,* vol. 32, no. 7, pp. 1239-1258, issue. 0162-8828, 2010.

[4] R. Labayrade, D. Aubert and J.P. Tarel “Real Time Obstacle Detection in Stereovision on non-flat road Geometry through V-disparity Representation,” *IEEE Intelligent Vehicle Symposium,* vol. 2, pp. 646-651, 2002.

[5] C.D. Jones, A.B. Smith, and E.F. Roberts, “A complete UV-disparity study for stereovision based 3D driving environment analysis,” *3-D Digital Imaging and Modeling, Fifth International Conference*, IEEE, pp. 204-211, 2005.

[6] W.T. Freeman, E.H. Adelson, “The Design and Use of Steerable Filters,” *Pattern Analysis and Machine Intelligence, IEEE,* vol. 13, no. 9, pp. 891-906, issue. 0162-8828, 2002.

[7] F.Chiabrando, R.Chiabrando, D.Piatti and F.Rinaudo, “Sensors for 3D Imaging: Metric Evaluation and Calibration of a CCD/CMOS Time-of-Flight Camera” *Pattern Analysis and Machine Intelligence, IEEE,* vol. 13, no. 9, pp. 891-906, issue. 0162-8828, 2002.

[8] R. Hartley and A. Zisserman, “Multiple view geometry in computer vision” Cambridge University Press, Cambridge, UK.