FORWARD STEREO OBSTACLE DETECTION WITH WEIGHTED HOUGH TRANSFORM AND LOCAL TEMPORAL CORRELATION

Feng Guo, Ling Cai, Ying Lin, and Rongrong Ji
{betop,lcai,linying,rrji}@xmu.edu.cn
School of Information Science and Engineering, Xiamen University

ABSTRACT

In this paper, we propose a robust obstacle detection approach by leveraging Weighted Hough Transform (WHT) in combination with temporal information correlation from the stereo video sequences. First, to model the road surface or obstacles in the video, rather than using simple threshold from binarized frame sequence, we propose to adopt WHT to extract the linear relation from the v-disparity map. Second, we exploit the temporal correlation of the road surface profile in stereo video sequences to produce the stable road surface model by filtering out the unreliable road. Extensive experiments with comparisons to state-of-the-arts demonstrate the merit of the proposed approach.

Index Terms— *Obstacle Detection, Weighted Hough Transform, Temporal Correlation, Stereo Vision*

1. INTRODUCTION

Coming with the popularity of automobiles, the autonomous driving system, a.k.a. Advanced Driver Assistance Systems (ADASs), Intelligent Vehicles (IVs) and Intelligent Transportation Systems (ITSs), has attracted extensive research interest and is providing great assistances to our daily driving tasks. However, it poses several challenges hesitating its proliferation, among which one of the most significant ones refers to the difficulty of reliable obstacle and road detection [1][2]. That is due to the complexity of the road environment, in which scenario none single method can handle various changes in road appearance and background/foreground occlusions.

Vision based technique serves as an important solution towards reliable obstacle and road detection. In this case, two calibrated cameras are equipped to produce the so-called v-disparity map. In this map, the line with inverse slope corresponds to the road surface, and the vertical line represents the obstacle in front of the cameras. And the task is to extract the inverse slope as well as the vertical line as robust and efficient as possible. To this end, a typical way is to use Hough transform to extract the linear relation, which requires a binarization operation on this v-disparity map with a pre-set threshold. Unfortunately, the threshold is difficult to determine and adaptive. In this paper, our first contribution is to tackle this issue by proposing a Weighted Hough Transform (WHT), which can reliably extract the linear relation from the v-disparity map without image binarization.

On the other hand, image sequences are often shaking when the cameras are moving, which inspire us to exploit the temporal information to assist us in obtaining the stable road profile. In this paper, our second contribution is to take advantage of such information to model the temporal correlation of road surface to provide the stable road profile.

By combining both steps, we come up with a very robust algorithm towards forward stereo obstacle detection. We test our approach on the widely-used KITTI[3][4] Vision Benchmark Suite and demonstrate the superior performance over several existing approaches and the state-of-the-arts. In terms of the speed, it requires 0.3 seconds in processing one frame, while in terms of the accuracy the overall performance is 85.3% in KITTI's urban marked road dataset.

This paper is organized as follows. Section 2 reviews some state-of-the-art approaches for obstacle detection. Section 3 specially focuses on the extraction of road surface feature by Weighted Hough Transform. Section 4 describes how the local temporal correlation helps to estimate the piecewise linear curve. The experimental results are represented in Section 5. Finally, Section 6 draws the conclusion.

2. RELATED WORK

Obstacle detection has drawn the attention of many computer vision researchers. In 2002, v-disparity image [5] is firstly proposed by *Labayrade* for obstacle detection. This method extracted the striking straight lines from the disparity maps to represent objects with same disparities. Similarly, the u-disparity map [6] can be obtained by

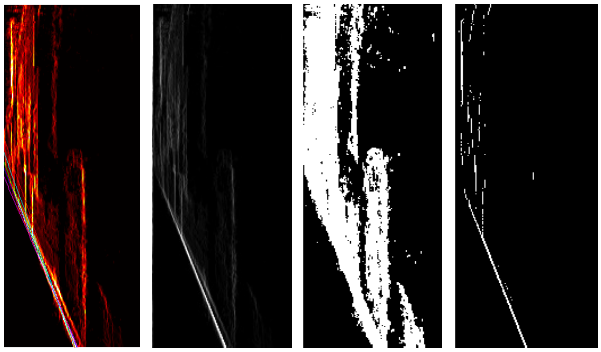
accumulating the pixels which have the same disparity and lie on a particular image row. It reveals the width of the obstacle in the image. Both two representation ways are based on the disparity maps, in image space where the resolution dramatically decreases with distance increasing. To tackle this issue, some improved approaches are proposed to use the measurements in the real-world coordinates, instead of the ones in the image space, to get a well-defined depth resolution [7]. The 3-D Euclidean sparse space has been directly used by *Alonso* et al [8]. In 2009, *Badino and Franke* proposed the stixel world [9] to produce a compact medium level representation for the 3-D world, so that the ground plane and the distance to the main objects in the scene can be estimated with the sticks rising from the ground.

3. ROAD SURFACE EXTRACTION WITH WEIGHTED HOUGH TRANSFORM

In most conventional approaches based on the v-disparity images, it is assumed that the linear model can approximately represent the road profiles in the disparity maps. We denote the v-disparity map F being the function imposed on the disparity image I_Δ, i.e., $F(I_\Delta)=I_{v\Delta}$. F is an accumulation map which counts the pixels that have the same disparity and lie on a given image line. For the line i on F, the abscissa corresponds to the different pixels with certain disparities. The strength indicates the number of pixels with the disparity Δ_M. This v-disparity method makes scene understanding possible, so that the ground plane and the 3-D obstacles can also be clearly recognized. Until now, the v-disparity map is widely used in road modeling and object detection in the context of intelligent vehicles.

To generate the accurate road surface profile on the v-disparity map, the image binarization has to be performed prior to filtering out the pixels with low frequency values. By this means, the environmental noise can be reduced greatly. However, how to determine the threshold value for image binarization is a key issue; it exerts a significant effect on the segmentation of free space and obstacles. In most cases, an inappropriate thesholding strategy will cause the serious misleading estimation to the road surface and the unsatisfactory detection result [10]. In this paper, we introduce the temporal weighted Hough Transform method working on v-disparity image binarization to overcome the difficulty of threshold value selection.

The conventional HT [11] was applied on the binary images to extract some specific geometry elements, but it is very hard to obtain the exact straight line. The intensity



(a) (b) (c) (d)

Fig. 1 Comparison of road surface extraction in v-disparity image by using (a) our approach (five straight lines are extracted) (b) image adjusting, (c) binarization, (d) most frequent value in each row.

of the noise or the improper threshold values during the binarization can easily mislead into generating the unsatisfactory results. The reason is that the threshold value determines the number of retained pixels, which is directly related to the computational cost of HT and the detection performance. A small threshold will retain the most edge pixels, but the noisy clutter will lead to the ambiguities in the peak selection. On the contrary, a high threshold will discard many details in the image, such as the small or far obstacle in front of cameras. To overcome the difficulty of threshold selection, we implement the weighted Hough Transform [12] in our approach so that we do not need to select threshold to binarize the disparity images. In conventional HT, the accumulation value from the binary image is calculated as:

$$Accumulator(\rho,\theta) = Accumulator(\rho,\theta) + F(\rho,\theta,x,y)$$

where $F(\rho,\theta,x,y)$ is defined as

$$F(\rho,\theta,x,y)=\begin{cases}1, & \rho = x\cos\theta + y\sin\theta,\ and\ I_\Delta'(x,y)=1\\ 0, & otherwise\end{cases}. \quad (1)$$

In formula (1), I_Δ' is the binarization of disparity image I_Δ, ρ and θ are accumulating parameters in HT. To detect the road profile far away from the camera, we take the low intensity information into consideration in the plane projection. We re-define F as

$$F'(\rho,\theta,x,y)=\begin{cases}kD(x,y), & \rho = x\cos\theta + y\sin\theta,\\ 0, & otherwise\end{cases} \quad (2)$$

where $D(x,y)$ is the intensity of the disparity in image corresponding to a pixel, and k is a factor to uniform the result to ensure $kD(x,y)$ is between 0 and 1. Fig. 1 displays the comparison among four methods that extract road surface in v-disparity image. Our approach can retain as

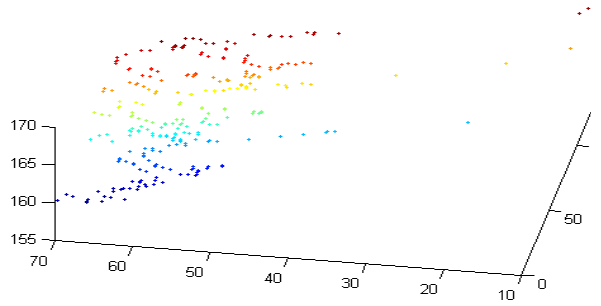


Fig. 2 Parameter pairs (ρ,θ) are locally correlated in a sample video sequence (containing140 frames)

many details as possible to process the far objects. Meanwhile, it could hardly be affected by noise. As shown in Fig. 1, all straight lines representing the road surface are extracted by our approach**.**

We model the road surface as a succession of parts of planes. In v-disparity map, it is projected as the piecewise linear curves for the following segmentation of free space and obstacles.

4. LOCAL TEMPORAL CORRELATION

Many approaches also utilize the temporal information to improve the final accuracy and remove the noise. In [13], the occupancy grid map has been used to filter out the false positive results from the disparity image. Kalman filter can generate an approximate B-spline road in [14].

To achieve the high-efficient process, we impose a constraint on the local correlation of road surface in our approach. As shown in Fig. 2, the parameter pair ρ and θ is locally correlated in the sequential frames, but the outliers are located far away from the local temporal "centroid". This correlation implies the temporal continuity of the road feature, and can be attributed to the stable traffic within a certain temporal range. In a smooth urban scene, most flat streets without wild fluctuation can be approximated as the linear curves, which results in the slight changes in the temporal domain. But in the case of complex conditions, the fluctuation of roads in the real world will result in the rapidly shaking road profiles in the v-disparity map, so that the temporal information of ρ and θ are invalid. With the temporal information, we can achieve the good result on the smooth urban roads.

Another advantage of the local temporal correlation is that outliers can be easily removed because of their salience. In most cases, the outliers in ρ-θ space can be attributed to detect large structural obstacles surfaces that dominate the disparity image. The final road surface profile curve S should change gradually with the conversion of the scene. Let ρ, θ be the expected parameters for the linear relationship in the v-disparity, and S be the expected piecewise linear curve that represents the road surface profile. We can define the optimization function $I(\rho, \theta, S)$ as

$$I = \arg\min_{\rho,\theta}\left(\sum_i \alpha_i\left(\left|\bar{\rho} - \rho_i\right| + \left|\bar{\theta} - \theta_i\right|\right) + \left|S_i - S_{i-1}\right|\right) \tag{3}$$

The first term is the difference of ρ and θ to their average values within the local time window, and the second term is the change of road profiles in the two adjacent frames, which aims at restraining the final $I(\rho, \theta, S)$ at a small scale.

In the video sequences, the parameter pair (μ_ρ, σ_ρ) and $(\mu_\theta, \sigma_\theta)$ can be updated by the statistical analysis on the previous result from our approach. The parmeter $w_{i,j}$ means the different weight of the j_{th} parameter in the frame i, and it represents the dominant degree of straight line in the road profile and the similarity in the time domain. The more dominant the linear relationship, the higher weight it get. The longer the frame happens, the lower the weight. Finally, if the piecewise linear curve S is out of the range, we will remove the linear relation with the small weight to minimize the $I(\rho, \theta, S)$.

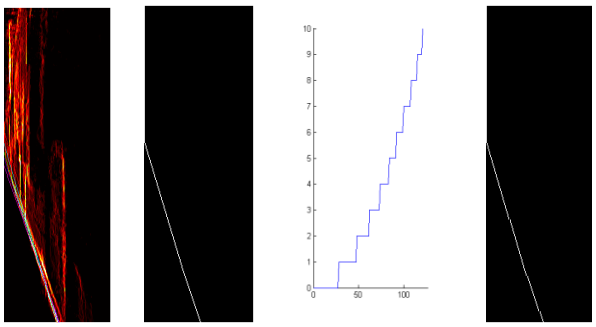
$$\begin{cases} \left|\rho - \mu_\rho\right| \le k\sigma_\rho \ \text{ and } \ \rho_{\min} \le \rho \le \rho_{\max} \\ \left|\theta - \mu_\theta\right| \le k\sigma_\theta \ \text{ and } \ \theta_{\min} \le \theta \le \theta_{\max} \end{cases} \tag{4}$$

$$\begin{cases} \mu_\rho = \dfrac{1}{N}\sum_i \pi_i \sum_j w_{ij}\rho_{ij} \\ \mu_\theta = \dfrac{1}{N}\sum_i \pi_i \sum_j w_{ij}\theta_{ij} \end{cases} \tag{5}$$

In the situation of the severe occlusion, such as traffic jam or a red light at a road across, the conventional road extraction is unreliable. However, the temporal local correlation can be used to forecast the road surface in the process of obstacle detection. The current road feature is compared with the proper road surface parameter occurring in the past short period time with weighted average, the parameters at the newer frames have the higher weights while the ones at the older frame has the lower weight. The result of comparison can check whether the road is smooth or not.

5. EXPERIMENTAL RESULTS

In our experiments, the gray scale stereo image pairs come from KITTI Vision Benchmark Suite. Two significant reasons make us to choose this system: firstly, KITTI benchmark can provide the latest and most



(a) (b) (c) (d)

Fig. 3 Example of road surface extraction in our proposed algorithm: (a) extract straight lines by WHT-VD, (b) generate piecewise linear curve that represents road surface model, (c) the tolerance factor corresponding to the disparity value, (d) final road profile curve with combination of tolerance factor.

challenging stereo video sequences from the real world with the detailed intrinsic and extrinsic parameters; secondly, the data from more than three sensors are described with the detailed parameters, which pave the way for further research work. The camera images are cropped to a size of 1382 x 512 pixels using libdc's format 7 mode. After rectification, the images get slightly smaller. The cameras are triggered at 10 frames per second in order to work with other devices synchronously. The final images we use to perform our experiment have 1242 x 375 pixels. Our approach is implemented in both Matlab and C++ with the library OpenCV, and it is run on a PC equipped with an Intel 2.9 GHz i5 CPU.

We compare many stereo matching algorithms on the KITTI dataset and find out which one are the most suits for the real traffic environment. Finally we decide to retrieval disparity images by Semi-global matching which compromises the speed from local methods and the accuracy from global methods. Here we choose the OpenCV implementation that uses Birchfield-Tomas cost function [5].

Fig. 3 shows the road surface extraction result from a frame. Note that more than 5 straight lines with parameters are extracted in our approach, and then a piecewise linear curve made up by the lines above, finally the robust criterion is generated in combination of the piecewise linear curve and the tolerance factor.

Fig. 4 shows the result when we use the local temporal correlation on sequence of disparity images. After the vehicle turns left, the road surface extraction does not work well at first and some lower obstacles cannot be detected properly, but these false negatives are corrected soon in the following frames when the road profiles are updated.

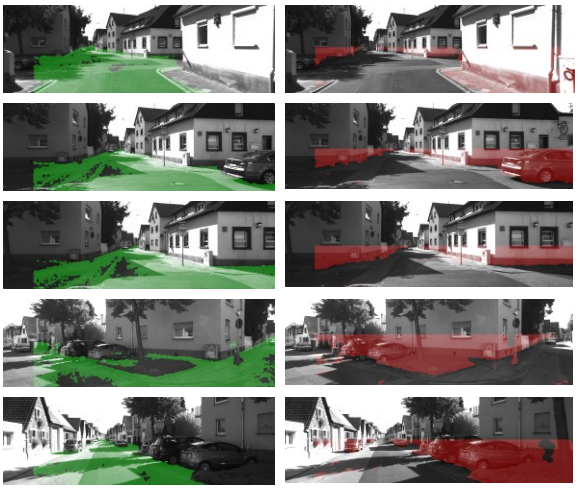


Fig .4 Free space(left column) and obstacle detection result(right column)

We use the classical pixel-based evaluation to evaluate road and lane estimation performance in the bird's-eye-view space, and the established measures come from [15]. We choose KITTI's urban marked road dataset, and our approach can process one frame in 0.3 second averagely, the precision is 85.3% and the recall is 86.5%.

6. CONCLUSION

In this paper, we present a robust method to perform obstacle detection in video sequences with WHT and local temporal correlation, which provides reliable and accurate obstacles and free space segmentation. To overcome the information loss in the road profile extraction, WHT is introduced to find the principal linear components in v-disparity domain. Besides, the local temporal correlation plays an important role of filtering out the wrong straight line parameter in some complicated scenes. The experimental results show that it successfully handles most cases of urban streets, rural highway etc. In future work, we will optimize the performance further by fusing more information from the additional sensors. Collision avoidance alert and obstacle segmentation will be significant directions in the following research.

7. ACKNOWLEDGEMENT

This work is supported by the Nature Science Foundation of China (No. 61422210 and No. 61373076), the Natural Science Foundation of Fujian Province of China (No.2014J01249), the Fundamental Research Funds for the Central Universities (No.2013121026), and the 985 project of Xiamen University.

8. REFERENCES

[1] A.B. Hillel, R.Lerner, D.Levi, et al. Recent progress in road and lane detection: a survey. Machine Vision and Applications, pages 1-19,2012.

[2] N.E.Faouzi, H.Leung , A.Kurian. Data fusion in intelligent transportation systems: Progress and challenges–A survey. Information Fusion, 12(1): 4-10, 2011.

[3] Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? The KITTI vision benchmark suite[C]//Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012: 3354-3361.

[4] Geiger A, Lenz P, Stiller C, et al. Vision meets robotics: The KITTI dataset[J]. The International Journal of Robotics Research, 2013: 0278364913491297.

[5] R. Labayrade, D. Aubert, J.-P. Tarel. Real time obstacle detection in stereovision on non flat road geometry through, Intelligent Vehicle Symposium, IEEE. pages 646-651, 202.

[6] Z. Hu, F. Lamosa, K. Uchimura. A complete uv-disparity study for stereovision based 3d driving environment analysis. 3-D Digital Imaging and Modeling (3DIM) 2005. Fifth International Conference on, IEEE, 2005, pages 204-211,2005.

[7] D. Llorca, M. Sotelo, I. Parra, J.E. Naranjo, M. Gavilán, S. Álvarez, An experimental study on pitch compensation in pedestrian-protection systems for collision avoidance and mitigation, Publisher, City, 2009.

[8] I.P. Alonso, D.F. Llorca, M.Á. Sotelo, L.M. Bergasa, P.R. de Toro, J. Nuevo, M. Ocaña, M.G. Garrido. Combination of feature extraction methods for SVM pedestrian detection, Publisher, City, 2007.

[9] H. Badino, U. Franke, D. Pfeiffer, The stixel world-a compact medium level representation of the 3d-world, in: Pattern Recognition, Springer, pages 51-60, 2009.

[10] N. Soquet, D. Aubert, N. Hautiere, Road segmentation supervised by an extended v-disparity algorithm for autonomous navigation. Intelligent Vehicles Symposium, IEEE, pages. 160-165, 2007.

[11] Hough, P.V.C.: Method and means for recognizing complex patterns. U.S. Patent 3,069,654,1962.

[12] M.K. Ibrahim, E. Ngau, M.F. Daemi, Weighted Hough transform. Intelligent Robots and Computer Vision X: Algorithms and Techniques, International Society for Optics and Photonics pages 237-241, 1992.

[13] K. Kohara, N. Suganuma, T. Negishi, T. Nanri. Obstacle detection based on occupancy grid maps using stereovision system, Publisher, City, 2010.

[14] A. Wedel, H. Badino, C. Rabe, H. Loose, U. Franke, D. Cremers, B-spline modeling of road surfaces with an application to free-space estimation, Publisher, City, 2009.

[15] J. Fritsch, T. Kuhnl, A. Geiger. A new performance measure and evaluation benchmark for road detection algorithms[C]//International Conference on Intelligent Transportation Systems (ITSC). 2013, 28: 38-61.