Non-flat Road Detection Based on A Local Descriptor

Kangru Wang, Lei Qu, Lili Chen, Yuzhang Gu, Dongchen Zhu, Xiaolin Zhang

Shanghai Institute of Microsystem and Information Technology (SIMIT), Chinese Academy of Sciences (CAS)

Abstrct **The detection of road surface and free space remains challenging for non-flat plane, especially with the varying latitudinal and longitudinal slope or in the case of multi-ground plane. In this paper, we propose a framework of the road surface detection with stereo vision. The main contribution of this paper is a newly proposed descriptor which is implemented in the disparity image to obtain a disparity feature image. The road regions can be distinguished from their surroundings effectively in the disparity feature image. Because the descriptor is implemented in the local area of the image, it can address well the problem of non-flat plane. And we also present a complete framework to detect the road surface regions base on the disparity feature image with convolutional neural network architecture.**

Index Terms **Road Detection, Free Space Detection, Local Descriptor, Stereo Vision, Convolutional Network.**

I.INTRODUCTION

Road surface and free space detection is a key component of intelligent vehicle applications and mobile robot system. The information of drivable space and understanding of the environment can improve traffic safety and efficiency. With the development of Driver Assistance Systems (ADAS) and Collision Avoidance Systems (CAS), frees pace detection has become an important research area of computer vision. Road surface estimation can be used for pitch angle compensation[[1] and improving the accuracy of obstacle detection [2], and free space estimation. Free space estimation is applied widely in many applications such as vehicle navigation [3] and pedestrian or vehicle detection [4, 5].

Labayrade et at[6] proposed the well-known ‘V-disparity algorithm’ which is a common approach for the road surface modeling. The algorithm simplifies the extraction of the 3D ground and obstacle into a 2D linear process without using any prior knowledge of the scene appearance. The ’V-disparity’ widely used to detect the road surface [7-12]. [8]use a u-disparity map to get the detail information of ground. Traditional methods based on V-disparity map have some limitations in detecting non-flat ground especially in the off-road environment. Some more robust algorithms have been proposed to extract the non-flat roads .In[6] the authors assume that the road can be modeled as piecewise linear and non-flat roads can be modeled by a few linear lines. [8, 13-15] address the problems with the road of different longitudinal slope. But these methods usually fail in complex scene especially with wide variance of latitudinal slope and multi-ground plane. [18] use sliding window paradigm to address the detection of the road with variable latitudinal slope and multi-ground plane. The plane is considered locally plane in very corresponding window, and sub-V-disparity map is created to represent the details of ground plane. But the number of window is hard to decide.

There are also some algorithms preserve the physical properties of the road in Euclidian space[17-21]. [16]estimates road surface using multivariate polynomial in the YZ plane domain .In[19,20],the input 3D map is reduced to a 2D map by accumulating all the points into a histogram of height versus distance which is similar to the v-disparity map creation. [21-23] applied a 2D quadratic surface fitting. [22, 23]introduces a method to estimate the planar patches for the Euclidian domain from the disparity map and then exploited the estimated patch parameters for eliminating outliers during road fitting. The traditional estimation of free space is usually based on the construction of occupancy grids [24, 25].The occupancy grid method models the occupancy evidence of the environment using a two-dimensional array or grid. Each cell of the occupancy grid maintains the probability of occupancy. However, these methods require the knowledge of the stereo sensor characteristics to compute depth map or 3D Euclidian point cloud as an initial step. With the development of deep learning, some algorithms to detect road or freespace have been proposed. [26] proposed a network to detect road that takes advantage of a large contextual window and uses a Network-in-Network (NiN) [27] proposed a multi-layer CNN architecture with a new loss function to detect free space.

Figure 1: the left images are stereo left images and the right images are the disparity feature images.

In this paper, we propose a framework of the road and free space estimation with stereo vision. The main contributions of this paper are presented as follows. First, we proposed a descriptor to obtain a feature map where the road regions can be distinguished from their surroundings effectively. Second, the feature map is segmented into superpixel regions and use a convolutional neural network architecture to classify every superpixel region. We use the contextual information around the consider superpixel region to improve the accuracy of the model.

II PROPOSED APPROACH

Our proposed method mainly consists of three steps, i.e. compute a disparity feature image, segment the disparity feature image, detect the road surface region using convolutional neural network architecture.

A. Compute the disparity feature image

In this paper, we propose a descriptor which can extract road and non-road surface feature from disparity image. A disparity feature image can be obtained after computer every pixel in disparity image with the descriptor. This feature of the disparity feature image can distinguish road regions from their surroundings effectively. This descriptor is implemented on the disparity image directly without using any other information. In this paper, the dense disparity estimation is performed using the algorithm of [28] with the reason of high quality.

In ideal conditions, the pixels on the road surface should have constant disparity in the same row, while the disparity value decreases gradually from the lower row to the upper row. Whereas all of the pixels of Frontal-parallel object surface should have the constant disparity. What's more, the pixels of vertical structures parallel to the road shoud have the constant disparity along the vertical coordinate while disparity was decrease progressively or increase progressively along the horizon coordinate.

Based on the disparity character of road and non-road surface, we define a descriptor for each pixel with a rectangular structure of 3x3 block, as show in Figure.2. We encode the center pixel in the center block b_0 by binary pattern. The descriptor is defined by comparing the average disparity of a sub region in the structure with other sub region. In this way, we can obtain a binary sequence which then is translated to a decimal value. The output value of the descriptor (i.e. the value of center pixel) can be obtained as follows:

$$O = \sum_{i=1}^{8} c_i 2^i$$

where $c_i (i = 0, \ldots 8)$ is defined as follow:

$$c_1 = \begin{cases} 1 & if\,(b_1 + b_2 + b_3)/3 < (b_4 + b_0 + b_5)/3 \\ 0 & else \end{cases} \quad c_2 = \begin{cases} 1 & if\,(b_4 + b_0 + b_5)/3 < (b_6 + b_7 + b_8)/3 \\ 0 & else \end{cases}$$

$$c_3 = \begin{cases} 1 & b_0 > b_2 \\ 0 & else \end{cases} \quad c_4 = \begin{cases} 1 & b_0 < b_7 \\ 0 & else \end{cases}$$

$$c_5 = \begin{cases} 1 & b_0 > b_1 \\ 0 & else \end{cases} \quad c_6 = \begin{cases} 1 & b_0 > b_3 \\ 0 & else \end{cases}$$

$$c_7 = \begin{cases} 1 & b_0 < b_6 \\ 0 & else \end{cases} \quad c_8 = \begin{cases} 1 & b_0 < b_8 \\ 0 & else \end{cases}$$

Where b_i is the average disparity value of the block region, as show in Figure.2.

b_1	b_2	b_3
b_4	b_0	b_5
b_6	b_7	b_8

Figure 2:the structure of the descriptor

Because the descriptor is implemented in the local area of the image, it can address the problem of non-flat plane. Fig.1 shows the example of the disparity feature image with different terrains. In the example of Fig.1, the size of each block of the descriptor is 1x1. The size of block can be changed according to the environment scene and the quality of disparity image.

B. Segment the disparity feature image

The second step in our proposed framework is to segment the disparity feature map with the SLIC superpixel algorithm. Simple linear iterative clustering (SLIC) [29] is a good algorithm to generate superpixels by using spatial and color information with computational and memory efficiency. Ideally, the small region segmented by superpixel belongs to the same object. In addition, it shows excellent boundary adherence which can help improve the precision of segment between road and non-road region.

Thus, we utilize the SLIC superpixel algorithm to segment the stereo left image into superpixel regions. Then, rule of segment is used to divide the disparity feature image into corresponding superpixel regions. The feature extracted for each superpixel is used to determine the region class (road or non-road). We consider all the pixels in the superpixel region have the same class.

C. detect road region

In this step, we propose a convolutional neural network architecture to classify every superpixel region into road or non-road. It consists of extracting patches around superpixel regions of the disparity feature image and predicting the label of the superpixel using a trained CNN.

To reduce the impact of disparity estimation error and improve the accuracy of prediction, a possible solution is to make use of the contextual information around the considered superpixel region. Thus, in this framework we input the image patches centered at the centroid of each superpixel to our network and the output is a class (road or non-road) of the considered superpixel region.

Because the feature in disparity feature image is very obvious, we use a simple net architecture inspired by the LeNet-5 architecture which is also proposed to implement on gray image initially. The CN architecture is listed in Table I.

TABLE I

Convolutional Neural Network architecture

	Type of layer	parameter
1	Convolutional layer	5x5x20
2	Non-linear	Relu
5	maximum	2x2
6	Convolutional layer	3x3x20
7	Non-linear	Relu
10	maximum	2x2
11	Fully connected layer	O=500
12	Non-linear	Relu
13	Fully connected layer	O=2

III.EXPERIMENT RESULT

A. Dataset

For the estimation experiments we use the raw data available in the KITTI dataset [31]. The dataset consists of 122 images as training images and 303 as testing images, and ground truth is generated by manually. We choose the training images from 7 diverse sequences (09_26_d_09, 09_26_d_18, 09_26_d_35, 09_26_d_48, 09_28_d_34, 09_28_d_45, 09_28_d_68). and the testing images are chosen from other 10 diverse sequences (09_26_d_17, 09_26_d_39, 09_26_d_61, 09_26_d_64, 09_26_d_86, 09_26_d_93, 09_28_d_66, 09_30_d_33, 10_03_d_27, 10_03_d_47).This diverse sequences are obtained from different categories including city, residential, road, person and campus where terrains rang from plane to non-plane ground.

B. Training Scheme

We created training and validation samples from the training images. Each sample consists of the image patch and its referent class. To create the samples, we extracted the image patch centered at the centroid of each superpixel region and the class (road or non-road) is attributed to the considered superpixel region. To make the balance of the samples, we sampled the non-road class samples. We conduct the training using stochastic gradient descent .

C. Evaluation Result

We compare our approach with two baselines: V-disparity [6] and Sub-V-disparity [18]. Fig.3 shows a sample of the obtained results. These demonstrate that our algorithm is able to provide superior average performance on non-flat ground. We also tested the effect of the size of block in our proposed descriptor on the model. The size of each block is 1x1, 3x3. A visual sample of the results can be seen in Fig. 4.

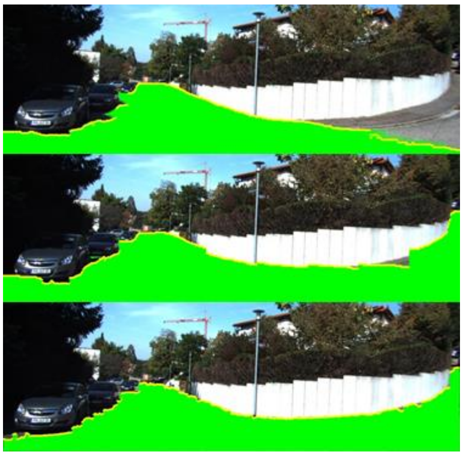

Figure 3: From the first to third column are the example results of v-disparity, sub-v-disparity and our method.

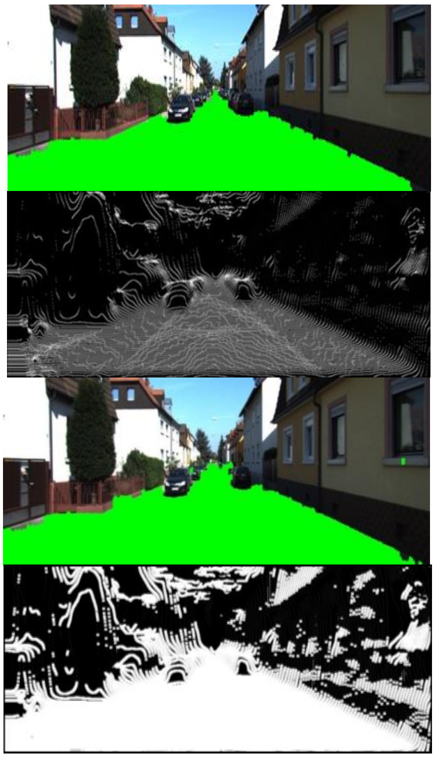

Figure 4: The images in the first and third columns are the road surface detection results with the descriptors of 1x1 block size and 3x3 block size. The images in the second and fourth columns are the disparity feature images obtained by the descriptors of 1x1 block size and 3x3 block size .

IV. CONCLUSION

We present a method of road and free space estimation with stereo vision. We proposed a descriptor to obtain a disparity feature map where the road regions can be distinguished from their surroundings effectively. A complete framework is proposed to detect the road surface region base on the disparity feature image with a convolutional neural network architecture. The framework is shown to provide robust results over a variety of terrains from KITTI's benchmark. Our framework also benefits traditional methods with better results.

REFERENCES

[1] R. Labayrade, D. Aubert, and J.-P. Tarel, "Real Time Obstacle Detection in Stereovision on Non Flat Road Geometry Through v-disparity Representation," *Intelligent Vehicle Symposium*, pp. 646-651, 2002.

[2] A. Wedel, H.Badino, C. Rabe *et al.*, "B-Spline Modeling of Road Surfaces With an Application to Free-Space Estimation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 4, pp. 572-583, 2009.

[3] M. Muffert, T. Milbich, D. Pfeiffer *et al.*, "May I Enter the Roundabout? A Time-To-Contact Computation Based on Stereo-Vision," *Intelligent Vehicles Symposium*, pp. 565-570, 2012.

[4] J. K. Suhr, K. M. Kang, and H. G. Jung, "Dense stereo-based critical area detection for active pedestrian protection system," *Electronics Letters*, vol. 48, no. 19, pp. 1199-1201, 2012.

[5] M. Enzweiler, M. Hummel, D. Pfeiffer *et al.*, "Efficient Stixel-Based Object Recognition," *Intelligent Vehicles Symposium*, 2012.

[6] R. Labayrade, D. Aubert, and J. P. Tarel, "Real Time Obstacle Detection in Stereovision on Non Flat Road Geometry Through "V-disparity"Representation," *Intelligent Vehicle Symposium*, 2002.

[7] R. Labayrade, and D. Aubert, "A Single Framework for Vehicle Roll, Pitch, Yaw Estimation and Obstacles Detection by Stereovision," *Intelligent Vehicles Symposium*, 2003.

[8] C. D.Jones, A.B.Smith, and E.F.Roberts, "A Complete U-V-Disparity Study for Stereovision Based 3D Driving Environment Analysis," *3-D Digital Imaging and Modeling, IEEE Fifth International Conference*, pp. 204-211, 2005.

[9] N. Soquet, D. Aubert, and N. Hautiere, "Road Segmentation Supervised by an Extended V-disparity Algorithm for Autonomous Navigation," *Intelligent Vehicles Symposium*, 2007.

[10] C. Teoh, C. Tan, and Y. C. Tan, "Ground Plane Detection for Autonomous Vehicle in Rainforest Terrain," *Sustainable Utilization and Development in Engineering and Technology*, 2010.

[11] J. Qin, and A. Ping, "Depth Extraction Method Based on Multi-view Stereo Matching," *Computer Engineering*, vol. 36, no. 14, pp. 174-176, 2010.

[12] K. Atsuta, and K. Hamamot, "A Robust Road Profile Estimation Method for Low Texture Stereo Images," *ICIP*, pp. 4273-4276, 2009.

[13] R. Labayrade, and D. Aubert, "A Single Framework for Vehicle Roll, Pitch, Yaw Estimation and Obstacles Detection by Stereovision," *Intelligent Vehicles Symposium*, pp. 31-36, 2003.

[14] N. Soquet, D. Aubert, and N. Hautiere, "Road Segmentation Supervised by an Extended V-disparity Algorithm for Autonomous Navigation," *Intelligent Vehicles Symposium*, pp. 160-165, 2007.

[15] C. W. Teoh, C. S. Tan, and Y. C. Tan, "Ground Plane Detection for Autonomous Vehicle in Rainforest Terrain," *Conference on Sustainable Utilization and Development in Engineering and Technology*, pp. 7-12, 2010.

[16] S. Nedevschi, R. Danescu, D. Frentiu *et al.*, "High accuracy stereovision approach for obstacle detection on non-planar roads," *In Proceedings of the Intelligent Engineering Systmes*, pp. 211–216, 2004.

[17] A. D. Sappa, F. Dornaika, D. Ponsa *et al.*, "An Efficient Approach to Onboard Stereo Vision System Pose Estimation," *Transactions on Intelligent Transportation Systems* vol. 9, no. 3, pp. 476 - 490, 2008.

[18] D. Yiruo, W. Wenjia, and K. Yukihiro, "Complex Ground Plane Detection Based on V-disparity Map in Off-road Environment," *Intelligent Vehicles Symposium* pp. 1137-1142, 2013.

[19] S. Nedevschi, R. Schmidt, T. Graf *et al.*, "3D lane detection system based on stereovision," *Intelligent Transportation Systems*, 2004.

[20] A. Wedel, H. n. Badino, C. Rabe *et al.*, "B-Spline Modeling of Road Surfaces With an Application to Free-Space Estimation " *Transactions on Intelligent Transportation Systems*, vol. 10, no. 4, pp. 572-583, 2009.

[21] F. Oniga, and S. Nedevschi, "Processing Dense Stereo Data Using Elevation Maps: Road Surface, Traffic Isle, and Obstacle Detection " *Transactions on Vehicular Technology* vol. 59, no. 3, pp. 1172 - 1182, 2010.

[22] X. Ai, Y. Gao, J. G. Rarity *et al.*, "Obstacle detection using U-disparity on quadratic road surfaces," *Intelligent Transportation Systems*, pp. 1352-1357, 2013.

[23] U. Ozgunalp, X. Ai, and N. Dahnoun, "Stereo vision-based road estimation assisted by efficient planar patch calculation," *Signal, Image and Video Processing*, vol. 10, no. 6, pp. 1127-1134, 2016.

[24] H. Badino, U. Franke, and D. Pfeiffer, "The stixel world-a compact medium level representation of the 3d-world," *Joint Pattern Recognition Symposium*, pp. 51-60, 2009.

[25] H. Badino, U. Franke, and R. Mester, "Free Space Computation Using Stochastic Occupancy Grids and Dynamic Programming," *Workshop on Dynamical Vision*, vol. 20, 2007.

[26] C. C. T. Mendes, V. Frémont, and D. F. Wolf, "Exploiting Fully Convolutional Neural Networks for Fast Road Detection," *International Conference on Robotics and Automation*, 2016.

[27] D. Levi, N. Garnett, and E. Fetaya, "StixelNet: A Deep Convolutional Network for Obstacle Detection and Road Segmentation," *BMVC*, 2015.

[28] Z. J, and L. Y, "Stereo matching by training a convolutional neural network to compare image patches," *Journal of Machine Learning Research*, vol. 17, pp. 1-32, 2016.

[29] R. Achanta, A. Shaji, K. Smith *et al.*, "SLIC Superpixels Compared to State-of-the-Art Superpixel Methods," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, pp. 2274-2282, 2012.

[30] M. Lin, Q. Chen, and S. Yan, "Network in network," *arXiv preprint arXiv,* vol. 1312, no. 4400, 2013.

[31] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," Computer Vision and Pattern Recognition (CVPR), vol. 2012 IEEE Conference on, no. IEEE, 2012.