CrossMark

Road Surface Segmentation based on Vertically Local Disparity Histogram for Stereo Camera

Shinji Kakegawa[1] · **Haruki Matono**[2] · **Hideaki Kido**[3] · **Takeshi Shima**[3]

© The Author(s) 2017. This article is published with open access at Springerlink.com

Abstract In this paper, we propose a road surface segmentation technique which is accurate and suitable for hardware implementation. The road surface is segmented by detecting the boundary between road and obstacle, based on the disparity histogram which we define as VLDH (Vertically Local Disparity Histogram). On each pixel of disparity image, VLDH is computed from the disparities of vertically local neighbourhood pixels. The major advantage is the feasibility of the pipeline processing on image processing hardware for stereo camera. The direct advantage on processing time is confirmed based on implementation into FPGA. Experimental result also shows that the accuracy of the proposed method is better than the conventional method.

Keywords VLDH · Disparity histogram · Road surface segmentation · Stereo camera · FPGA

1 Introduction

Environment perception around vehicle is essential for the realization of the ADAS (Advanced Driver Assistance Systems) functions and for the autonomous driving function. Although maps and communication with infrastructure give an a priori knowledge about the environment, autonomously sensing the environment using on-vehicle sensors is indispensable in the natural dynamic situations.

From on-vehicle sensors, various obstacles are detected. Typical targets are the road users such as vehicle and pedestrian as well as the structure which makes the road boundary such as wall and guardrail. The small obstacles on the road such as curb stone, speed bump, pot hole, are challenging but also demanded targets. The recognition of the free space, the drivable space free from obstacle for ego-vehicle, is also a demanded capability to the on-vehicle sensor for autonomous driving.

The proper recognition of the road surface is important and can be the base technology for the obstacle detection and free space recognition. Since obstacle can be defined as a structure standing on the road, it is beneficial to use the road profile information as a geometric constraint for reliable detection of obstacle.

In this paper, we focus on the road surface segmentation technique for the stereo camera. Stereo camera is the sensor that delivers 3-D information as well as 2-D information of the scene by outputting the disparity (distance) image and visual image. Road surface segmentation explained in this paper is the technique to extract the road surface region in the 2-D image. This information can be used for the estimation of road profile by referring the 3-D information (disparity image) inside the segmented road surface region.

Road surface segmentation has an important role for the accurate estimation of the road profile. A classical approach of the road profile estimation is presented in [1]. The road profile is estimated by detecting a line in the V-disparity histogram space. V-disparity histogram is the disparity histogram on each image row. V-disparity histogram can be represented as 2-D image, V-disparity image. In V-disparity image, one axis is the image row v, the other axis is the disparity, and each

✉ Shinji Kakegawa
shinji.kakegawa@hitachi-automotive.eu

[1] Hitachi Automotive Systems Europe GmbH, Engineering Center, 18 rue Grange Dame Rose, 78140 Velizy, France

[2] Hitachi Automotive Systems Ltd, Technology Development Division, Hitachinaka, Japan

[3] Hitachi Ltd, Hitachi Research Laboratory, Hitachi, Japan

grid contains the histogram count. On the assumption of the planar road without bank angle, the occupied location of the road surface becomes aligned in the V-disparity histogram space. Therefore, the road profile can be estimated by fitting a line on V-disparity image. In [1], the input for the road profile estimation is the whole disparity image and the segmentation of the road surface is not conducted. In the cluttered environment where the road surface is not majority in the image, however, it is known that the result can be erroneous. For accurate and robust estimation, it is important to segment the road surface and input only the disparities of road surface to the road profile estimation [2, 3].

In addition to the accuracy aspect, an important requirement for on-vehicle stereo camera is the real-time processing in high frame rate. Since the CPU resource is limited and the hardware processing can usually run much faster than CPU processing, the computationally expensive process is preferred to be implemented in hardware as much as possible.

On hardware processing, to minimize the latency and achieve the fast processing, it is often important that every processing block is performed in pipeline [5]. On the image processing hardware for stereo camera, since the captured images stream in raster order, it is common that the operations for stereo processing such as left and right image acquisition, lens distortion correction, rectification, and disparity computation can be all processed in pipeline, and finally the disparity image is output in raster order to be processed further by CPU.

On this background, we developed a road surface segmentation technique which is accurate and suitable for the hardware implementation with pipeline processing on stereo camera hardware platform.

2 Review of the Conventional Approach

There exist several segmentation methods to extract road surface using disparity image obtained by stereo camera [2–4]. The authors consider the U-disparity approach applied in [2] is a simple and effective approach for the purpose of the accurate road profile estimation, and regard as the base algorithm for improvement.

U-disparity histogram is the disparity histogram in each column and its histogram count can be represented as

$$UD\left(u,\overline{d}\right) = \sum_{j=1}^{m} \delta\left(D(u,j),\overline{d}\right) \tag{1}$$

UD is defined per each image column and per each quantized disparity value $\overline{d}$. (u,v) is the image vertical and horizontal location respectively, D(u,v) is the disparity at (u,v) from disparity image, and m is the image height. δ is the Kronecker delta which returns 1 when the disparity D(u,v) is contained in the quantized disparity $\overline{d}$ and returns 0 otherwise.

There is a characteristic that UD(u, D(u,v)) becomes large when obstacle exists at disparity D(u,v). The intuitive explanation is as follows. Assume the U-disparity histogram value at certain disparity is large. Then it is suggested there are many pixels which have same disparity (distance) on the image column, which in turn suggests there exists an obstacle at the distance since it is assumed that obstacle is standing close vertically to the camera to have same disparity on the obstacle surface.

Using this characteristic, the road surface can be segmented by extracting the pixels (u,v) whose corresponding histogram count UD(u, D(u,v)) is small [2].

There is a concern in this approach, however, that the false extraction of the road surface is expected at the close horizontal surfaces such as the top surface of the road shoulder or other obstacle, since the road surface is just segmented as 'non-obstacle' pixel, or 'non-vertical' surface. To obtain more reasonable area of the road surface, one idea is to detect the boundary location between road and obstacle in the image. U-disparity histogram, however, is not a suitable representation for the boundary detection since the vertical location information is lost as seen in Eq. (1).

Another shortage is the low feasibility for the hardware implementation on stereo camera. The main reason is that the U-disparity histogram is computed by reading all the vertical pixel values on the disparity image. Since disparity on each pixel is usually computed in raster order in the image processing hardware, the road surface segmentation process has to be waited until the whole disparity image is computed. This nature makes the pipeline processing almost impossible and the hardware implementation become difficult because the latency and the local memory consumption tends to increase too much.

To solve these issues, a new representation of disparity histogram is introduced in the next section.

3 Vertically Local Disparity Histogram

Vertically Local Disparity Histogram (VLDH) is defined as a disparity histogram which is computed from the disparities of vertically local neighborhood of a reference pixel. Its histogram count can be represented as

$$\text{VLDH}\left(u,v,\overline{d}\right) = \sum_{j=0}^{N-1} \delta\left(D(u,v+j),\overline{d}\right) \tag{2}$$

VLDH is defined per each image location (u,v) and per each quantized disparity value $\overline{d}$. N denotes the pixel range of the 'neighborhood'. N is a parameter, but generally it should be constituted by a small number of pixels to be able to sense enough the obstacle existence.
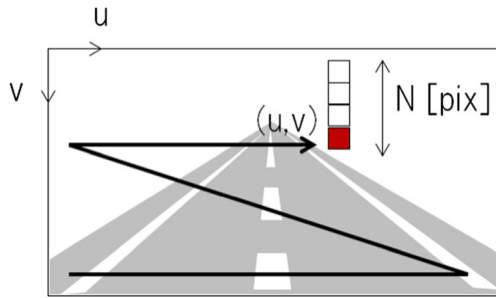
Fig. 1 Schematic image of VLDH computation with N pixels window scanned in raster order
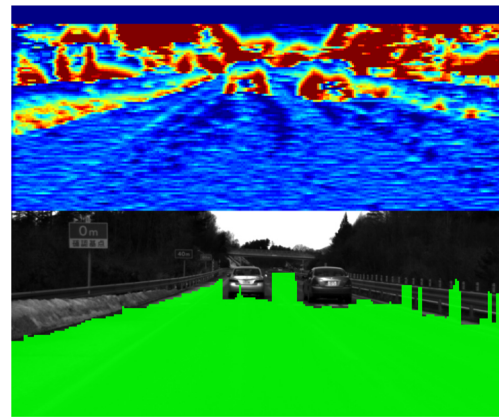

Fig. 2 c(u,v) map (*top*) and segmentation result of road surface (*bottom*)

There is a characteristic that VLDH(u,v,D(u,v)) becomes large when obstacle exists at disparity D(u,v) and at around neighborhood of (u,v).

The important difference with U-disparity histogram is that the vertical location information is preserved. This characteristic makes it possible to detect the boundary between road and obstacle as detailed in the next section.

Another difference is the input characteristics on the computation. U-disparity histogram is computed from the vertical all pixels per image column. On the other hand, VLDH is computed from the vertical neighbourhood pixels per image pixel. This local nature on input is preferable for the hardware implementation as discussed in the next section.

4 Road Surface Segmentation Using VLDH

We take an approach to detect the boundary between obstacle and road surface, and segment pixels under the boundary as road. The boundary location can be defined reasonably as the foot position of the nearest obstacle in each image column. VLDH becomes a suitable representation in order to detect the existence of obstacle and the image location of its foot position at the same time.

In essence, we detect the boundary at the lowest image location where VLDH(u,v,D(u,v)) is large enough. More precisely, we compute the following histogram count,

$$c(u, v) = \sum_{\substack{u_i \in [u - \Delta u, \ u + \Delta u] \\ v_i \in [v - \Delta v, \ v + \Delta v] \\ \Delta d_i \in [-\Delta d, \ +\Delta d]}} \mathrm{VLDH}(u_i, v_i, D(u_i, v_i) + \Delta d_i) \quad (3)$$

which is the filtered value of the VLDH(u,v,D(u,v)), and Δu, Δv, Δd are the filtering parameters. Δu, Δv can be adjusted so as to suppress the noisy measurement of the boundary position and smooth the boundary position in horizontal direction, by adjusting the window area to compute VLDH feature. Δd, which determines the disparity variation tolerance on obstacle surface, can be adjusted based on the possible disparity measurement error by stereo camera. Then, assuming that the pixel-wise processing is performed in bottom-up raster order as shown in Fig. 1, the boundary is detected at the image location where threshold judgment c(u,v) > cth holds in the first time per each column u.

However, VLDH is represented as 3-D array and the memory efficiency is not good if Eq. (3) is implemented as it is defined. In actual implementation, it is not necessarily to preserve VLDH, and c(u,v) can be directly computed by counting

Table 1 Pseudo code of the computation of c(u,v)

```
Pseudo code
c(u,v) = 0
for u_i = u - Δu: u + Δu
    for v_i = v - Δv: v + Δv
        for v_n = v_i – (N-1): v_i
            If d(u_i,v_n) > = d(u_i,v_i) – Δd & d(u_i,v_n) < = d(u_i,v_i) + Δd
                c(u,v) = c(u,v) + 1
            end
        end
    end
end
```
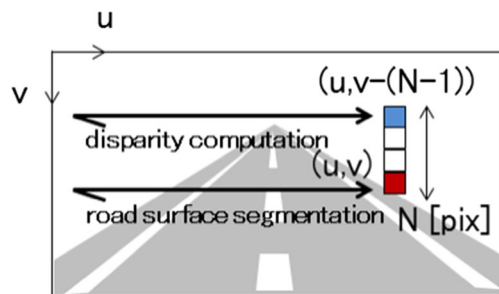

Fig. 3 Image of the pipeline processing of the road surface segmentation and the disparity computation with N lines timing difference

Table 2 Conditions and parameters of segmentation

Stereo camera	Baseline 160[mm]
Disparity image	H240 x V100[pix] Generated by ZSAD block matching [6]
Proposed method parameters	$N = 10$[pix], $c_{th} = 17$[−] , $\Delta u = 2$[pix], $\Delta v = 0$[pix], $\Delta d = 0.375$[pix]

through the neighborhood N pixels on each pixel. This alternative and simpler form of the computation of $c(u,v)$ is shown in Table 1 as a pseudo code.

Figure 2 shows an example of the $c(u,v)$ map and the segmentation result of road surface. $c(u,v)$ map is the visualization of the $c(u,v)$ in each pixel location on the image, and warmer color represents the larger counts on the pixel location. From the $c(u,v)$ map, it is noticed that the histogram count is mostly large where obstacles exist. From the segmentation result, it is noticed that the road surface is well segmented over the free space up to the left and right guardrails and the preceding vehicles.

The proposed method is suitable for hardware implementation on stereo camera. The segmentation can be processed in raster order and it reads only the local N pixels in vertical direction on its each pixel-wise processing. Owing to this local nature on input, the process can be pipelined with disparity computation only with about N lines processing delay with about N lines line buffer, as shown in Fig. 3. Furthermore, this implementation should not add obvious latency because the processing time, or latency, is basically only affected by the most time consuming process among the pipelined processes, which is most probably the disparity computation.

It is noted that the last (N-1) lines on the top of image cannot be processed for the lack of the N lines input for segmentation, as seen in Fig. 2 (top) where there is no output for several lines on the image top. This is trivial for the road surface segmentation since the road surface is not usually located on the top of image.
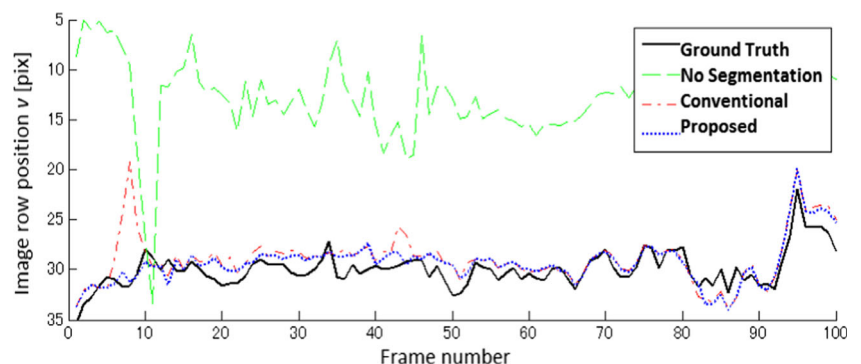
5 Evaluation on the Accuracy

5.1 Quantitative Evaluation on Crowded Scene

The accuracy of the road profile estimation is quantitatively evaluated, where the input of the road profile estimation comes from the following 3 set-up of road surface segmentation.

(1) No segmentation (whole area is regarded as road)
(2) Conventional method (U-disparity approach [2])
(3) Proposed method

The conditions and parameters of segmentation are shown in Table 2.

The evaluation scene is chosen on a crowded scene with obstacles in urban environment. The scene length is 100 frames, and each frame is captured each 0.5 [s]. The evaluation indicator is the error [pix] of the estimated road height at the 100 [m] ahead. The road height profile was estimated by fitting a line using the least square method to the disparities of the segmented road surface in V-disparity space. The ground truth of the road height profile is obtained by segmenting the road surface by visual judgment and fitting a line in V-disparity space similarly.

Result of the estimated road height for the evaluation scene is shown in Fig. 4, and the statistics on the road height error are shown in Table 3.

As seen in Fig. 4, the error of the road profile estimation using the proposed method is largely reduced compared to 'no segmentation' case, and mostly similar with the conventional method. Looking into Table 3 for the average error and the standard deviation of the error, it is confirmed that the proposed method gives slightly better result than the conventional method. The difference is mostly coming from the result at the 8th frame of the evaluated scene.

Figure 5 shows the segmentation result and the line fitting result on V-disparity image at the 8th frame. The upper side of the cargo truck on the right hand side of the image is falsely judged as the road surface for the conventional method. It is

Fig. 4 Estimated road height at 100 [m] ahead
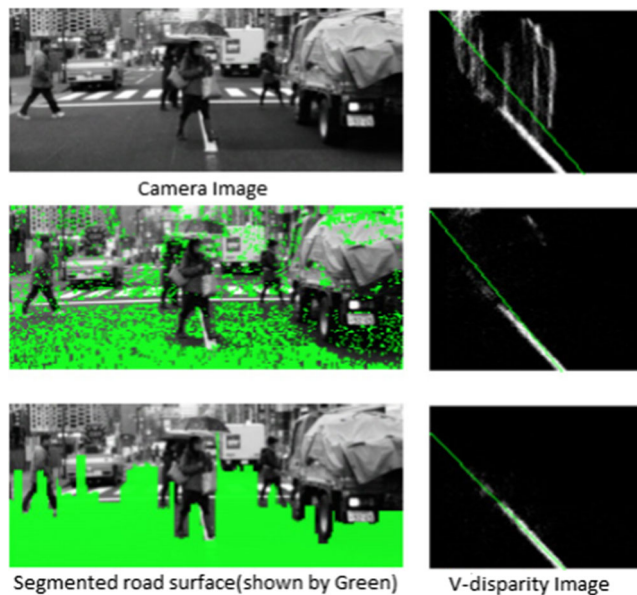
Table 3 Statistics on the road height error

	(1)No segm.	(2)Conven-tional	(3)Proposed
Average error[pix]	17.0	1.1	0.6
Standard deviation error[pix]	4.4	1.9	1.2

because the disparity is actually changing vertically in this region similar to the road surface. On the other hand, the segmentation is satisfactory for the proposed method since the proposed method segments road surface based on the detected boundary between road and obstacle, and the boundary location is well detected.
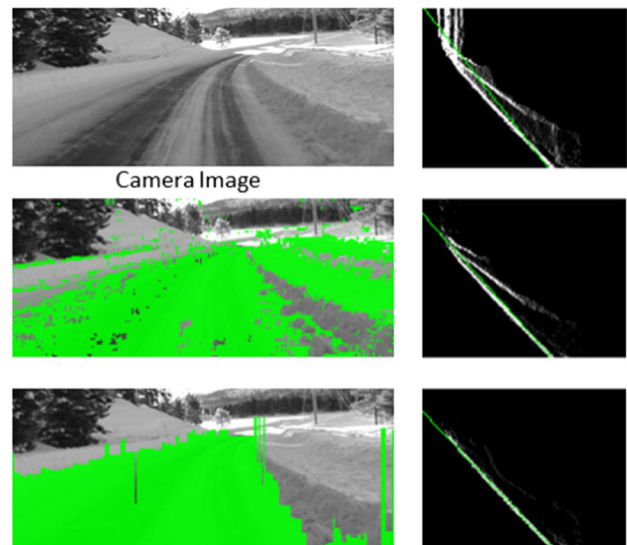
5.2 Qualitative Evaluation on Snowy Road

The segmentation performance between the conventional method and proposed method was further compared qualitatively by checking the segmentation result by human eye. The selected evaluation scene was a snowy road scene where the snow is mostly piled-up in left and right hand side of the road. The results of total 300 frames were inspected.

The result showed the clearer difference on the output between 2 methods. The proposed method generally showed the better segmentation outcome. A typical frame and its result are shown in Fig. 6. As seen in the figure, the top surface of the piled-up snow is falsely extracted as road surface for conventional method, whereas the segmentation of the boundary is
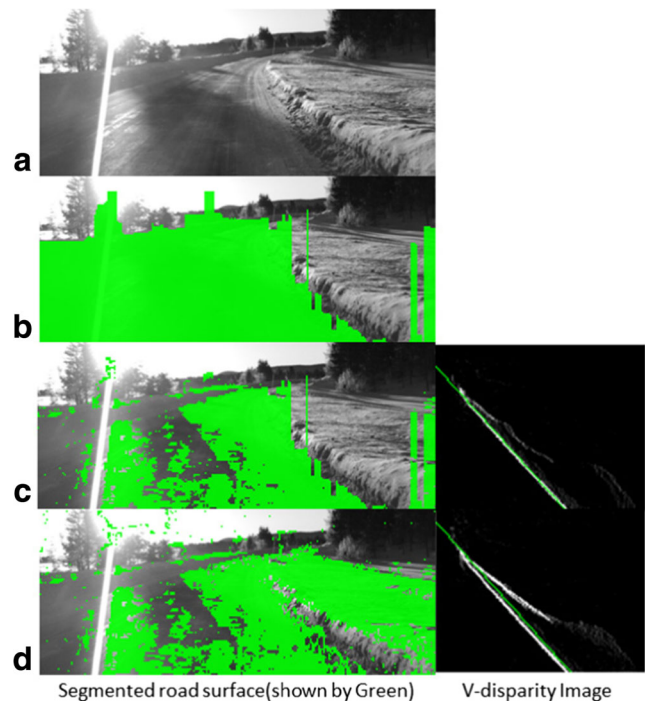

Fig. 6 Segmentation result and V-disparity image for 'no segmentation' (*top*), conventional (*mid*), and proposed (*bottom*) method

better for proposed method. The influence of this difference on the V-disparity image is clear. For conventional method, the disparities of many non-road pixels are input to the V-disparity image, which makes the line shape on the V-disparity image noisy and makes the line fitting difficult. It is noticed that the fitted line is estimated slightly upward than the actual profile for the conventional method, whereas the fitted line is accurate for the proposed method.

Fig. 5 Segmentation result and V-disparity image with fitted line for 'no segmentation' (*top*), conventional (*mid*), and proposed (*bottom*) method

Fig. 7 Segmentation result and V-disparity image. **a** Camera image. **b** Proposed method. **c** Proposed method (only showing the pixels of valid disparities). **d** Conventional method

Table 4 Processing time on PC

	Processing time[ms]
Conventional	13.8
Proposed	16.5

Figure 7 shows an example where the detected boundary position is not accurate for proposed method. The boundary detected by proposed method, (b) in Fig. 7, is not accurate globally in the left hand side, and locally in the right hand side at near region. The inaccuracy in the left hand side stems from the sparse input disparities on this region. The disparities in the low-textued region are not obtained in the prior disparity computation process, in the conditon where camera is facing against the sun. It is confirmed by (c) in Fig. 7 that this type of inaccuracy does not affect much since the actual pixels of valid disparities inside this area are mostly accurate and only these pixels are used for generation of V-disparity histogram. The inaccuracy in the right hand side in near region stems from that the nearest obstacle of snow wall is not visible since it is out of sensor field of view, so that the boundary was typically detected at the second nearest obstacle. It is noted that the segmentation result is nevertheless better than conventional method, comparing (c) and (d) in Fig. 7.

Overall, it is concluded that the proposed method shows better performance than the conventional method in our evaluation scenes.

6 Evaluation on the Processing Time

We evaluated the processing time of the road surface segmentation with the road profile estimation on a PC (Intel Xeon 2.8GHz) and on a general embedded system (ARM CPU with/without FPGA). The time was measured from the time of the disparity image output to the time of the road profile estimation output. The disparity image computation is performed in a separated image processing hardware which outputs the disparity data line by line. The conventional method is implemented in software, and the proposed method is implemented in 2 ways, one in software, and one in FPGA.

Table 5 Processing time on an embedded system for CPU processing or FPGA/CPU processing. (The processing time is more correctly described as the latency between the disparity image output and the road profile estimation output)

	Processing time[ms]
Conventional(CPU)	81.3
Proposed(CPU)	87.2
Proposed(FPGA/CPU)	1.0 (less than 0.05 ms for CPU processing)

Fig. 8 The segmented obstacle surface locating above the ground level which is estimated by conventional method (*top*) and proposed method (*bottom*) at the 8th frame of the evaluation scene

For the software implementation of both methods, it is noted that we did not optimize the codes for high speed processing. We coded the both algorithms in Matlab script, and used the C codes generated by Matlab auto code function for the evaluation.

For the FPGA implementation of proposed method, the major processing of the segmentation and V-disparity computation are processed in FPGA, and the subsequent line fitting was only processed by CPU. The implementation into FPGA was conducted so as to pipeline the road surface segmentation with the disparity image computation.

The evaluation results are shown in Table 4, and Table 5.

From the previous tables, it is seen that the proposed method tends to be more time consuming than the conventional method when it is implemented in software, but much faster when it is implemented into hardware. It is noted that the implementation of the proposed method into FPGA only adds the latency of 1.0 millisecond to the original hardware processing without segmentation, owing to the pipelining capability of the proposed method.

7 Influence to the Environment Perception

In Fig. 8, we show the result of the segmented obstacle surface on the same scene as Fig. 5. The obstacle surface was
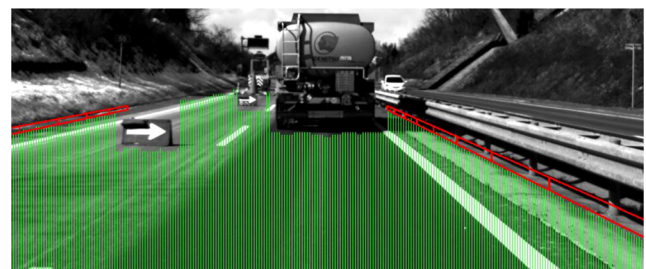
Fig. 9 Result of the road edge detection and free space recognition

extracted by simply segmenting the disparities which is above the estimated road height profile in 3-D space. From the figure, it can be seen that the proposed method is giving a better result than the conventional method, since the obstacle surface such as the pedestrian foot or the far vehicle is segmented correctly with proposed method.

This example indicates that the accurate road profile estimation helps for the detection of obstacle. This influence is expected to be more obvious for the more difficult detection target of small obstacles such as curb stone and pothole, or the obstacle in the far distance.

In Fig. 9, the result of the road edge detection and free space recognition is shown. The red line shows the road edge and green area shows the free space. Road edges and other obstacles on road were detected effectively using the geometric constraints in relation to the road height profile. The free space was defined as the region in front of the road edges and obstacles on road. Based on the accurate road profile, it is possible to detect small obstacles and far obstacles, to result in the good perception of the environment.

8 Conclusion

We proposed a road surface segmentation technique which is accurate and suitable for the hardware implementation in stereo camera platform by using a newly defined disparity histogram feature.

The suitability of the hardware implementation was confirmed by the implementation of the proposed method into FPGA. The evaluation of this implementation showed that it only adds a latency of 1.0 millisecond to the hardware processing on stereo camera, owing to the pipelining capability of the proposed method. Experimental result also showed that the accuracy of the proposed method is better than the conventional method.

The proposed road surface segmentation technique contributes on the accurate road profile estimation and obstacle detection for stereo camera. The interested future works are as follows:

(1) Road profile estimation using the surface model of higher degree of freedom to cope with the undulating road surface
(2) Road edge detection, Bump / Pot-hole detection for the precise definition of free space
(3) Road shape recognition such as the road curve, merge, fork, and crossroads for scene understanding

References

1. Labayrade, R., et al.: Real time obstacle detection on non flat road geometry through 'V-Disparity' representation. In: Proc. IEEE Intell. Veh. Symp., vol. **2**, pp. 646–651 (2002)
2. Musleh, B., et al.: U-V Disparity analysis in urban environments. In: Proc. the 13th Int. Conf. Comput. Aided Sys. Theory, vol. **2**, pp. 426–432 (2011)
3. Wedel, A., et al.: B-spline modeling of road surfaces with an application to free-space estimation. Proc. IEEE Trans. Intell. Transp. Syst. **10**, 572–583 (2009)
4. Badino, H., et al.: Free Space Computation Using Stochastic Occupancy Grids and Dynamic Programming. Workshop on Dynamical Vision, ICCV (2007)
5. Johnston, C.T., et al.: Implementing image processing algorithms on FPGAs. In: Proc. 11th Electronics New Zealand Conference, Palmerston North, pp. 118–123 (2004)
6. Hirschmuller, H., et al.: Evaluation of stereo matching costs on images with radiometric differences. IEEE Trans. Pattern Anal. Mach. Intell. **31**, 1582–1599 (2009)

Shinji Kakegawa received the M.Sc. in mechanical engineering in Kobe University, Japan, in 2010. In 2010, He joined Hitachi Research Laboratory, Hitachi Ltd., Japan. He worked for development on stereo matching, stereo calibration, and road surface estimation technology, for on-vehicle stereo camera. Since 2015, he has been with Engineering Center, Paris office, Hitachi Automotive Systems Europe GmbH, France. He is currently working for the development of road edge detection and free space recognition technology for on-vehicle stereo camera.

Haruki Matono received the M.Sc. degrees in electrical engineering in Tohoku University, Japan, in 2005. In 2005, He joined Consumer Electronics Laboratory, Hitachi Ltd., Japan. He was with Hitachi Research Laboratory, Hitachi Ltd., Japan, for two years. He worked for the development of distance measurement technology and hardware platform for on-vehicle stereo camera. Since 2014, he has been with Technology Development Division, Hitachi Automotive Systems Ltd., Japan. He is currently working for the development of on-vehicle stereo camera as embedded system designer.

Hideaki Kido received the M.Sc. degrees in Systems Life Sciences in Kyushu University, Japan, in 2007. In 2007, He joined Hitachi Research Laboratory, Hitachi Ltd., Japan. He worked for development of hardware accelerator for image processing and image recognition algorithm. He was with Technology Development Division, Hitachi Automotive Systems Ltd., Japan, for two years. He worked for development on front-vehicle detection technology for on-vehicle stereo camera. Since 2016, he has been with Hitachi Research Laboratory and continuing the technology development for on-vehicle stereo camera.

Takeshi Shima received the M.Sc. degrees in school of engineering in University of Tokyo, Japan, in 1996. In 1996, He joined Hitachi Research Laboratory, Hitachi Ltd., Japan. In 2007, he was with the Robotics Institute, Carnegie Mellon University, USA, as a visiting researcher for one year. He is currently the chief researcher in smart systems research department in Hitachi, Ltd. He has been worked for autonomous driving system, image recognition technology for on-vehicle camera, and stereo camera technology.