A ROBUST ROAD PROFILE ESTIMATION METHOD FOR LOW TEXTURE STEREO IMAGES

Le Thanh SACH, Kiyoaki ATSUTA, Kazuhiko HAMAMOTO, Shozo KONDO*

School of Information and Telecommunication Engineering
Department of Information Media Technology, Tokai University.

ABSTRACT

The estimation of road profiles from low texture stereo images is a problematic task because the disparity images computed from such class of images have a large number of noisy disparities. This paper presents a new method that is based on edge maps to guide the cost aggregation process in the stereo matching problem. Using the proposed aggregation method, the disparity images are smooth at low texture regions, but the boundaries of on-road objects are still preserved. The V-disparity images computed from such reliable disparity images can clearly show the road profiles. Thereby, the road profiles can be straightforwardly extracted by the dynamic programming technique. Experiments on a long and real stereo image sequence demonstrate that the proposed method can robustly estimate the road profiles. Furthermore, on-road objects can be detected by combining v- and u-disparity images as well because their boundaries are preserved in the disparity images.

***Index Terms*—** Road Profile Estimation, Stereo Matching, Dynamic Programming, Cost Aggregation

1. INTRODUCTION

Autonomous navigation is an active and useful research field, as demonstrated by the competitions organized by DARPA [1]. Recent researches show that the stereo-based approach has an enormous potential for the navigation problem because it can utilize much information contained in stereo images and can determine the relative distance from on-road objects to the host vehicle. In stereo-based navigation problem, road profile estimation is a preliminary and important task because it highly influences the accuracy of on-road object detection in the next step.

As explained in [2], road surfaces can be represented by a succession of parts of planes in disparity spaces. However, because of the lack of textures in the stereo images, the disparity images contain a majority of noisy disparities. Although there are many research works based on disparity images to estimate road surfaces, a robust method for low texture images is still missing.

Only reliable disparities of sparse disparity images computed for high gradient pixels were used in [2, 3]. Assuming that the stereo cameras have small roll angles, road profiles can be detected by using Hough-Transform in v-disparity images [2]. For more robustness against noisy and outlier disparities, road surfaces were estimated by RANSAC in [3]. In order to obtain better distribution of gradients, an adaptive thresholding method was proposed in [4].

Because only using the disparities at high gradient pixels is unreliable for low texture images, the reliable disparities are propagated into neighboring pixels of low texture regions by an algorithm so-called *"Quasi-Dense"* [5]. However, the low texture problem has not been solved completely.

Dense disparity images were used in [6, 7]. In order to deal with noisy disparities, some heuristic ways have been proposed to select the disparities belonging to road planes. After that, the road planes were estimated by Least-Squares [6] and IRLS [7]. However, the robustness of those methods seems to rely on predefined anchor points to which the road planes are tied.

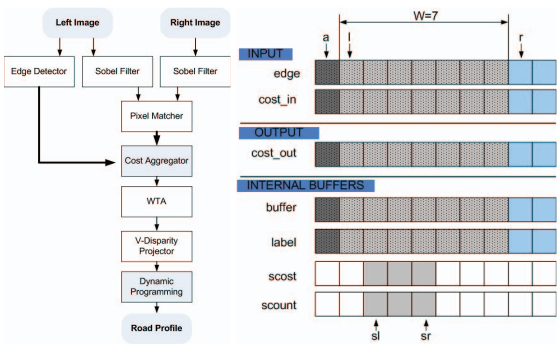
The matching of only phases of horizontal gradients was proposed in [8]. However, relying on matching costs of single pixels is unreliable enough for low texture images [9]. For more robustness, the aggregation of pixel matching costs in very wide windows was proposed in [10]. However, using the classical aggregation method with very wide windows will make disparity images smooth everywhere, i.e. on-road objects will disappear. Therefore, on-road objects can not be correctly detected. Moreover, as indicated in [10], both of the methods in [8] and [10] can not be used unless stereo cameras have small pitch angles.

This paper proposes a new method for the aggregation of matching costs. Using the proposed method, disparity surfaces are smooth at low texture regions, but the boundaries of on-road objects are still preserved. In addition, the computational complexity of the proposed method is independent to the size aggregation windows. Because there is a large number of reliable disparities belonging to road, road profiles clearly appear on v-disparity images and can be detected robustly by the dynamic programming technique.

*Email: LTSACH@cse.hcmut.edu.vn. I would like to thank Japan International Cooperation Agency (JICA) for funding my research

2. SYSTEM OVERVIEW

As shown in Fig. 1(a), input images are filtered by a horizontal Sobel kernel to emphasize textures, similar to [8]. After that, disparity images are computed by three modules: *Pixel Matcher*, *Cost Aggregator* and *WTA* (Winner-Takes-All).



(a) (b)

Fig. 1. (a) Road profile extraction system. (b) Data structures used for the proposed cost aggregation method.

Pixel Matcher computes a modified cosine distance defined by Eq. 1, where $\vec{g}_p = [g_p^r, g_p^g, g_p^b]^T$ and $\vec{g}_q = [g_q^r, g_q^g, g_q^b]^T$ are the results of Sobel filter for RGB components of pixel *p* and *q* respectively. L1-norm replaces L2-norm in the original cosine distance for a faster computation.

$$f(p,q) = 1 - \frac{g_p^r g_q^r + g_p^g g_q^g + g_p^b g_q^b}{(|g_p^r| + |g_p^g| + |g_p^b|)(|g_q^r| + |g_q^g| + |g_q^b|)} \simeq 1 - \cos(\vec{g}_p, \vec{g}_q) \qquad (1)$$

The output of *Pixel Matcher* is a 3D volume of matching costs denoted by *cost(u,v,d)*; where $U \times V$ is the size of image, D is the disparity search range, $u \in [0, U\text{-}1]$, $v \in [0, V\text{-}1]$ and $d \in [0, D\text{-}1]$. Different from the other methods, *Cost Aggregator* uses edge maps to guide the aggregation, as explained in the next section. Its output is denoted as $cost_{aggr}(u,v,d)$. *WTA* assigns disparity d_{best} to pixel $p(u,v)$ if $d_{best} = \arg\min_d cost_{aggr}(u,v,d)$.

Based the output of module *WTA*, V-disparity image which is formed by horizontally accumulating the image pixels with the same disparity is computed by *V-disparity projector*. Road profile is then extracted from the v-disparity image by module *Dynamic Programming* (DP).

3. COST AGGREGATION

The proposed cost aggregation method is inspired by the success of recent research works for stereo matching based on image segmentation [11]. However, different from the segmentation based methods that define a segment as a homogeneous region in 2D image, the proposed method defines a segment as a group of consecutive pixels on horizontal lines or vertical columns of the edge map computed for the reference image (the left image). For example, pixels from *s* to *(t-1)* in a line or a column of the edge map are grouped to a segment if and only if the line or the column changes value ($0 \rightarrow 1$ or $1 \rightarrow 0$) from *(s-1)* to *s* and from *(t-1)* to *t*, and there is no value change from *s* to *(t-1)*.

Table 1. Cost aggregation guided by edge maps

(a)	**Add new cost on the rightmost side of the window:**
1.	*buffer[r] = cost_in[r];*
2.	*aggr_cost = buffer[a] + cost_in[r];*
3.	**if**(edge[r-1] ^ edge[r]){ scost[++sr]=cost_in[r]; scount[sr]=1;}
4.	**else** { scost[sr]+= cost_in[r];scount[sr] +=1;}
5.	label[r] = sr;
(b)	**Remove the element that the window has left:**
6.	**if**(edge[r-W] ^ edge[r-W+1]) sl = sl +1;
7.	**else**{ scost[sl]-= buffer[l]; scount -=1;}
8.	*buffer[l] = aggr_cost - buffer[l];*
(c)	**Compute the average cost for the window:**
9.	mid = r - (W DIV 2); seg = label[mid];
10.	**if**((W-scount[seg]) != 0){
11.	cost_out[mid] = scost[seg]/scount[seg] + alpha*(buffer[l] - scost[seg])/(W-scount[seg]);
12.	} **else** cost_out[mid] = scost[seg]/scount[seg];
13.	a++; l++; r++

The proposed method performs the aggregation for horizontal lines and then for vertical columns of each U-V slice of *cost(u,v,d)*. Each line or column and the corresponding one in the edge map are represented by array *cost_in* and *edge* in Fig. 1(b) respectively. Aggregated costs will be written to *cost_out*. Similar to [12], *buffer[l]*$\rightarrow$ *buffer[r-1]* are used to store elements of the current window. The accumulated value (a summation) of those elements is saved to *buffer[a]*. *buffer[r]* is used to store the new incoming cost at *cost_in[r]* when the window moves to right.

A window may occupy several segments. These segments are indexed by *sl* $\rightarrow$ *sr*. The number of pixels in each segment and the accumulated cost of those pixels are stored to *scount* and *scost* respectively. Array *label* is used as a mapping from pixels to the segments to which the pixels belongs. The segment that contains the middle pixel of the current window is referred to as the *mid-segment* hereafter. Before processing *cost_in*, one initial segment is putted to *scount* and *scost*. This initialization is done by: (1) set *sl=sr=0*, (2) set *scost[0]*= the accumulated value of *W* (that is the length of windows) left border elements, (3) set *scount[0]=W*, and (4) set 0 (0: initial segment) to *label* for *W* left border elements.

The procedure to process each cost in *cost_in* is given in Table 1. In line (1) and (2), the proposed method saves the new cost *cost_in[r]* to *buffer[r]* and then adds this cost to the accumulated value of the previous window. A new segment will be created, in line (3), if there is a value change on the edge array at the rightmost element of the window. Otherwise, the new cost is accumulated to the rightmost segment identified by *sr*, in line (4). The mapping *label* is filled in line (5). The element that the window has left on the leftmost side, i.e *buffer[l]*, is removed in line (8). Before that, it is removed from the leftmost segment, in line (7), if the leftmost segment is still occupied by the window. Otherwise, the leftmost segment is removed by increasing *sl* by 1.

Different from the classical moving average [12], the proposed method reduces the influence of pixels that are outside of the *mid-segment* by multiplying a factor *alpha* (≤ 1) to the average cost of those pixels before adding this cost to the average cost of pixels inside the *mid-segment*, line (11).

4. ROAD PROFILE EXTRACTION

Given a disparity image, the intensity of a pixel *(v, d)* in a v-disparity image is the total number of the pixels in horizontal line *v* of the disparity image that have the disparity of *d*. V-disparity images are normalized by dividing the value of each pixel *(v, d)* by the maximum value in row *v*. The results are denoted by $I_{vd}(v, d)$, as shown in Fig. 2(d).

The slanted white line in Fig. 2(d) is the road profile. It is extracted by minimizing *Acc(v,d)*, defined by Eq. (2). Where $cost_{vd}(v, d) \stackrel{\text{def}}{=} e^{-I_{vd}(v,d)}$, and $\alpha(v)$ is the smooth factor for pixels in horizontal line *v* of $cost_{vd}(v, d)$. $\alpha(v)$ is selected as 10^{th} percentile of values in line *v* of $cost_{vd}(v, d)$. The minimum path, called $path_{DP}$, is found from the bottom to the top horizontal line of $I_{vd}(v, d)$.

$$Acc(v, d) = cost_{vd}(v, d) + \min \begin{pmatrix} Acc(v+1, d-1) + \alpha(v) \\ Acc(v+1, d) \\ Acc(v+1, d+1) + \alpha(v) \end{pmatrix} \quad (2)$$

From the bottom to the top horizontal line of a v-disparity image, $path_{DP}$ can turn to parallel to the vertical line *d=0* at $d = d_{inf}$ and $d = d_{obj}$, where d_{inf} is the disparity of objects at the infinity; for verged stereo cameras, $d_{inf} \neq 0$. d_{inf} can be determined at the calibration step. d_{obj} is the disparity of on-road objects that occlude a large area of the road. If there are no on-road objects then $d_{obj} \equiv d_{inf}$, as shown in Fig. 2(c). The extracted road profile in this paper is defined as a part of $path_{DP}$ from d=d_{obj} to d=D-1.

5. EXPERIMENTAL RESULTS

The proposed method is evaluated with a real image sequence from [6]. Two low texture images that were also used in [6, 10] are shown in Fig. 2(a) and Fig. 3(b). The disparity image computed by using the classical moving average (cost function: SAD, window size: 11×11, validation: left-right checking) for Fig. 2(a) is shown in Fig. 2(f). Because of low textures, almost all of disparities of road pixels are noisy. Therefore, v-disparity images computed by methods in [2, 4] are unreliable, as shown in Fig. 2(g-i). In contrast, the proposed aggregation method (edge detector: LoG with a threshold value of 0.002, window size:11×181, *alpha*: 0.2) can produce smooth disparity surface for the road area, as shown in Fig. 2(e). With the proposed aggregation method, road profiles can appear clearly in v-disparity images, as shown in Fig. 2(d). The red color path in Fig. 2(c) is the path discovered by the dynamic programming technique.

Because stereo images in the sequence were captured with flat road, road profiles can be estimated by least-squares method from the extracted ones yielded by DP, as shown in Fig. 2(b). Based on the results of the least-squares method, pitch angles of the stereo cameras can be computed using Eq. (10) in [2], as sketched in Fig. 3(e) for 864 stereo pairs in the sequence.

The most important feature of the proposed aggregation method is that the proposed method smooths disparities in low texture regions but preserves the boundary of on-road objects. As shown in Fig. 2(e) and Fig. 3(c), disparities belonging to car and electric pillars are still preserved. Thereby, on-road objects can be detected by combining v- and u-disparity images like [5], as shown in Fig. 3(b). In contrast, even the efficient cost function defined in Section 2 is used, the resultant disparity image will be over-smooth, as shown in Fig. 3(d), if it is computed by using the classical moving average with very wide window mentioned above.

6. CONCLUSION

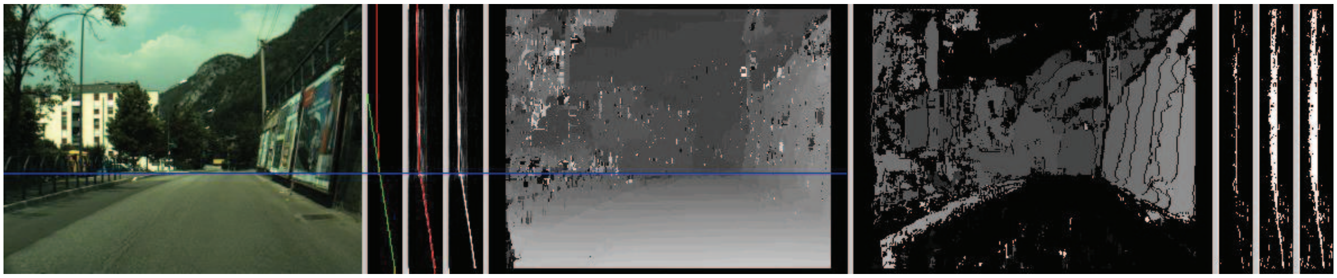
A new method named cost aggregation guided by edge maps is proposed in this paper. Like the classical moving average, its time complexity is independent to the size of aggregation windows. Particularly, the proposed aggregation method can smooth disparity images at low texture regions but preserve the boundary of on-road objects. Using the proposed method, road profiles can clearly appear in V-disparity images. They are robustly extracted by using the dynamic programming technique without any a priori knowledge about the ground plane like many other methods.

7. REFERENCES

[1] "www.grandchallenge.org: Defense advanced research projects agency (darpa)," .

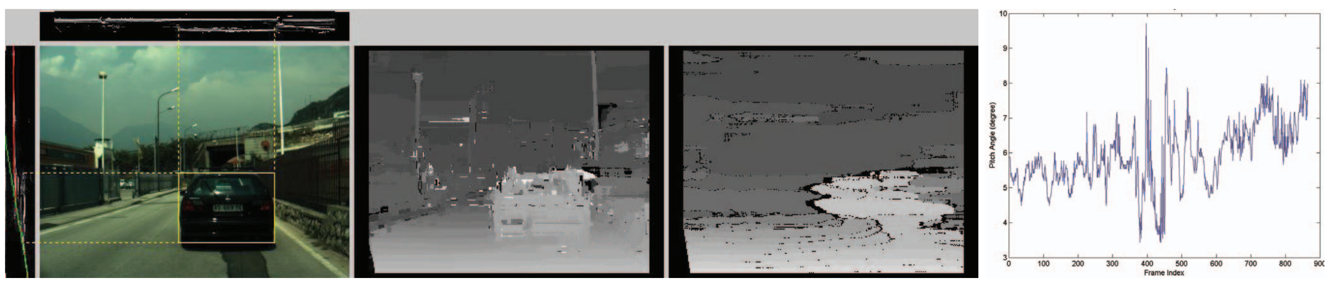
[2] D. Aubert R. Labayrade and J.P. Tarel, "Real time obstacle detection in stereo vision on non-flat road geometry through v-disparity representation," in *Proc. of Intelligent Vehicle Symposium*. IEEE, 2002, vol. 2, pp. 646–651.

[3] S. Se and M. Brady, "Ground plane estimation, error analysis and applications," in *Trans. on Robotics and*



(a) (b) (c) (d) (e) (f) (g) (h) (i)

Fig. 2. (a) The left image of a stereo pair. (b) Road profiles on v-disparity plane: $path_{DP}$ is in red color, green path is obtained by the least-squares method. (c) $path_{DP}$ (red color) is superimposed on the v-disparity image. (d) V-disparity image obtained by using the proposed aggregation method. (e) Disparity image computed by the proposed method. (f) Disparity image computed by the classical aggregation method. (g)-(i) V-disparity images computed by the adaptive threshold method [4] with 99%, 97% and 95% of high gradient pixels respectively.



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

Fig. 3. UP: u-disparity image. DOWN:(a) Road profiles on v-disparity image: $path_{DP}$ is in red color, green path is obtained by the least-squares method. (b) The left image of a stereo pair and the detected car (c) Disparity image computed by using the proposed aggregation method. (d) Disparity image computed by using classical moving average method for the same window size used for (c). (e) pitch angles (in degree) of stereo cameras computed from 864 stereo pairs in the sequence.

Autonomous Systems. ELSEVIER, 2002, vol. 39, pp. 59–71.

[4] D. Aubert N. Soquet and D. Hautiere, "Road segmentation supervised by an extended v-disparity algorithm for autonomous navigation," in *Proc. of Intelligent Vehicle Symposium*. IEEE, 2007, pp. 160–165.

[5] M. Perrollaz N. Hautiere, R. Labayrade and D. Aubert, "Road scene analysis by stereovision: a robust and quasi-dense approach," in *Proc. of ICARCV*. IEEE, 2006, pp. 1–6.

[6] Michele Zanin, "Localization of ahead vehicle with onboard stereo cemeras," in *Proc. of Image Analysis and Processing*. IEEE, 2007, pp. 111–116.

[7] Nikolay Chumerin and Marc M. Van Hulle, "Ground plane estimation based on dense stereo disparity," in *Proc. of Control, Automation, Robotics and Vision*. IEEE, 2006, pp. 1–6.

[8] R.I. Fedriga A. Broggi, C. Caraffi and P. Grisleri, "Obstacle detection with stereo vision for off-road vehicle navigation," in *Proc. of CVPR*. IEEE, 2005, pp. 65–72.

[9] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," in *International Journal of Computer Vision*, 2002, vol. 47(1/2/3), pp. 7–42.

[10] J. Katupitiya Z. Zhao and J. Ward, "Global correlation based ground plane estimation using v-disparity image," in *Proc. of Robotics and Automation*. IEEE, 2007, pp. 529–534.

[11] M. Gerrits and P. Bekaert, "Local stereo matching with segmentation-based outlier rejection," in *Proc. of Computer and Robot Vision*. IEEE, 2006, pp. 66–72.

[12] R. Hesser K. Muhlmann, D. Maier and R. Manner, "Calculating dense disparity maps from color stereo images, an efficient implementation," in *Proc. of Stereo and Multi-Baseline Vision*. IEEE, 2001, pp. 30–36.