Proceedings of the
2007 IEEE Intelligent Vehicles Symposium
Istanbul, Turkey, June 13-15, 2007

A Global Optimization Algorithm for Real-Time On-Board Stereo Obstacle Detection Systems

Susumu Kubota, Tsuyoshi Nakano, and Yasukazu Okamoto
Corporate Research & Development Center, TOSHIBA Corporation
Kawasaki 212-8582, Japan
{susumu.kubota, tsuyoshi4.nakano, yasukazu.okamoto}@toshiba.co.jp

***Abstract*— A fast and robust stereo algorithm for on-board obstacle detection systems is proposed. The proposed method finds the optimum road-obstacle boundary which provides the most consistent interpretation of the input stereo image pair. Global optimization combined with a robust matching measure enables stable detection of obstacles under various circumstances, such as heavy rain and severe lighting conditions. The processing time for VGA size image pair is about 15msec on a 3.6GHz pentium IV processor, which is fast enough for real-time applications.**

I. Introduction

Obstacle detection is one of the key elements of driver assistance systems for traffic safety and many vision-based methods have been proposed[10]. Because of availability of low cost cameras and ever increasing computational power of processors, stereo-based approaches are becoming promising.

General stereo algorithms have been extensively studied for decades among computer vision researchers[9]. Since on-board stereo systems have to process input images in real-time, it is desirable that computational cost of stereo algorithms is low. Besides, systems have to work stably under various conditions, such as heavy rain and severe lighting conditions. These two requirements make stereo-based approaches to obstacle detection very challenging.

In order to reduce computational cost, several methods that use inverse perspective mapping (IPM) have been proposed[4], [5], [8]. IPM is used to map the right image so that road regions in the remapped image match those in the left image under the flat road hypothesis[10]. Therefore, obstacles are detected by simply calculating the difference between the remapped image and the target image. However, those methods do not always work well under various conditions. For example, it is hard to avoid false detections when there are reflections on a wet road surface.

Several methods that use dense disparity map have also been proposed recently[6], [11]. Though they are computationally heavier than IPM-based methods since they require extensive correspondence search, it is getting feasible nowadays. However, computational cost is not the only problem.

First, it is difficult to find 1-to-1 correspondence where a region of interest only contains horizontal edges or there is no texture at all. This is called the aperture problem (see area A of Fig.1). The larger the aperture is, the less significant

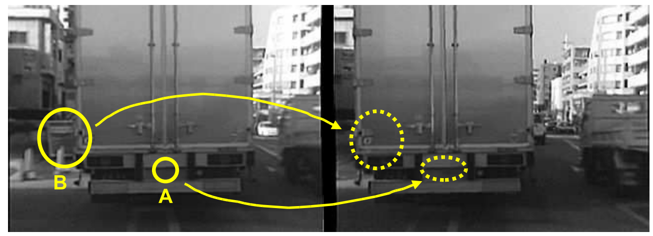

Fig. 1. Since Area A consists of horizontal line segments, it is difficult to find 1-to-1 correspondence. The signboard in area B is occluded by the truck and it is not visible in the right image.

the aperture problem becomes. However, using large aperture produces inaccurate depth estimation in the neighbors of depth discontinuities because of the violation of smoothness assumption which the most of the local correspondence search methods rely on.

Secondly, occlusions cause difficulties. For a half-occluded area, which is visible in one image but is not in the other one because of occlusion, there is no corresponding area (see area B of Fig.1). In order to avoid false detection at half-occluded areas, a stereo algorithm has to recognize half-occluded areas and exclude those areas from disparity computation.

Thirdly, the constant luminance assumption that many correspondence search methods rely on does not hold under severe lighting conditions (see Fig.2). Therefore, the distance measure that evaluate the goodness of a match has to be robust under practical conditions.

In this paper, we propose a global optimization algorithm for on-board stereo obstacle detection systems. Instead of locally solving correspondence problem, the proposed method gives a globally consistent interpretation of entire stereo image pair. Since the global optimization process explicitly handles depth discontinuities and occlusions, the aperture problem is alleviated while the accuracy of depth estimation around occlusion boundaries is maintained. The optimization is done efficiently with dynamic programming (DP) so that the proposed method can be used for real-time applications.

II. Basic Ideas

Fig.3 shows the configuration of our stereo camera system where the optical axises are parallel to each other and the baseline is parallel to the road surface. Cameras are assumed

Fig. 2. Because of severe lighting condition, constant luminance assumption does not hold in this pair of images

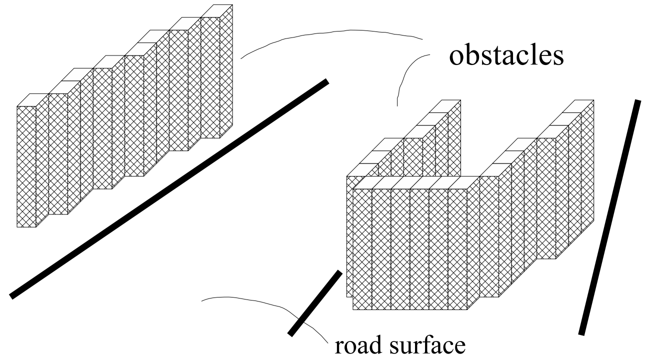

Fig. 4. Road environment model

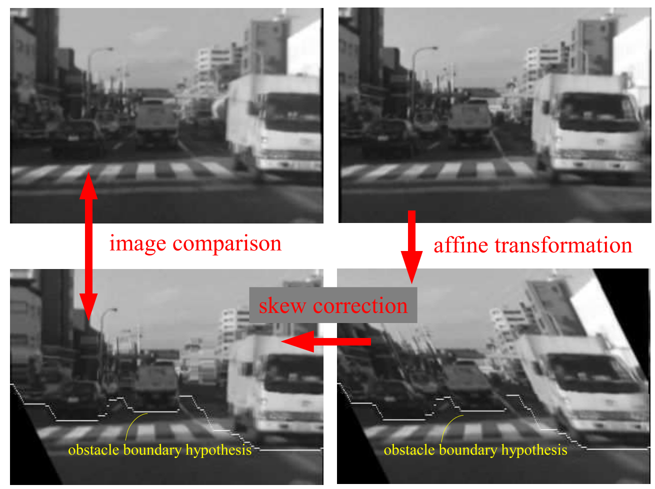

Fig. 5. Image remapping scheme

to be calibrated[12] and images are rectified[1] so that epipolar lines are parallel to horizontal scanlines of images.

The proposed method approximate the road environment with simple building blocks which assume that a road surface is flat and an obstacle consists of a set of columns that stand perpendicular to the road surface (see Fig.4). This road environment model is the one commonly used for post-processing of dense-stereo-based obstacle detection methods where pixels which have the same disparity and horizontal position are aggregated to form obstacle hypothesis[6], [11].

Since the model poses a constraint on the 3D structures of the road environment, the degree of freedom of the disparity map is significantly reduced. We refer to this constraint as the road environment constraint. Instead of using the constraint for postprocessing, the proposed method incorporate it to stereo calculation and derives a disparity map that best explains a stereo image pair.

Fig.5 explains the basic ideas of the proposed method. Top left and top right images in Fig.5 are rectified input stereo images. Suppose we know the parameters of road surface plane relative to the camera coordinate. The right image can be remapped so that the road area of the remapped image matches that of the left image (see bottom right image in Fig.5). Since the stereo image pair is rectified, this mapping is described with an affine transformation. Though the road areas in top left and bottom right images in Fig. 5 match well each other, the obstacles in the bottom right image in Fig.5 are skewed and do not match with those in the top left image. Then suppose we know the boundary between the road area and the obstacles (see white lines in the bottom right image in Fig. 5). With the boundary and the skew parameter which can be derived from the affine transformation parameters, we can correct the skew of the obstacle areas above the boundary so that the road and obstacles in the remapped right image match well with those of the left image (see top left and bottom left images in Fig.5). Though the bottom left image in Fig.5 is remapped from the original right image with the road plane parameters and road-obstacle boundaries, it matches well with the original left image.

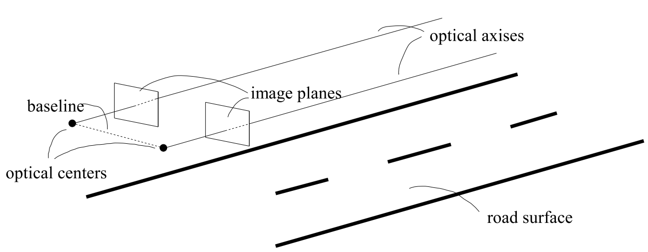

Fig. 3. Stereo camera system

The above example shows the following things:

1) Road plane parameters and road-obstacle boundary determine 1-to-1 correspondence between left and right image under the road environment constraint.
2) Given an appropriate boundary, the remapped right image matches well with the left image for not only road areas but also obstacle areas.

Instead of calculating 2D depth map, the stereo problem is now reduced to finding the boundary between road and obstacles that gives the best match between the left and right images under the road environment constraint, which is essentially a 1D optimization problem. An efficient algorithm using DP to solve this global optimization problem is proposed in the following section.

III. Algorithm Details

The basic procedure of the proposed method consists of the following steps:

1) Estimate road plane parameters.
2) Calculate a disparity space image (DSI).

3) Find the best path in the DSI which represent the boundary between the road and obstacles.

In order to handle a wide range of disparity and to reduce computational cost, multiresolution scheme can be applied where the basic procedure is applied from the coarsest level to the finest level.

A. Road Parameter Estimation

The road parameters can be estimated by using Labayrade's V-disparity technique[7].

Since images are rectified and the camera baseline is assumed to be parallel to the road plane, the projections of a point on the road surface to the left and the right images satisfy the following equation,

$$\begin{pmatrix} x_l \\ y_l \end{pmatrix} = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_r \\ y_r \end{pmatrix} + \begin{pmatrix} -bv_y \\ 0 \end{pmatrix}, \quad (1)$$

where (x_l, y_l) and (x_r, y_r) are the coordinates of the projected points in the left and right image respectively, v_y is y-coordinate of the vanishing line of the road plane, and b is a parameter concerning the width of the baseline and the camera height. We refer to (1) as the road plane constraint. This constraint gives stereo correspondence for the pixels in road areas. From (1), the relation between disparity value d and y_r is given as follows,

$$d = b(y_r - v_y), \quad (2)$$

where $d = x_l - x_r$.

Since relative position and orientation of the road plane are subject to change as the vehicle moves, the road plane constraint parameters (i.e. b and v_y in (1)) have to be estimated for each frame.

Fig.6 shows the edge pixels in the left image (white) and those of the remapped right image (black) where the default road plane constraint parameters are used for the mapping. The black pixels in the road area deviate with horizontal shift from the corresponding white pixels because of the inaccurate parameters. Eq.(1) shows that a linear functions with respect to y gives the amount of the deviation. Therefore, parameters can be estimated by calculating edge correlations for each scanline with respect to horizontal shift and applying Hough transform. In order to increase the stability of the estimation, the obstacle areas detected by the previous frame are masked out from the calculation of the edge correlations. Fig.7 shows an example of the remapping with updated parameters.

B. Disparity Space Image

The disparity space image (DSI) represents the matching score between the reference and the target images with respect to disparity and horizontal position[3]. The DSI has been used in DP-based scanline optimization methods where a pixel of the DSI represents a matching score between a pixel of a reference scanline and that of the target scanline. The proposed method modifies the DSI so that a pixel of the DSI represents a matching score for a column of pixels of the reference image under the road environment constraint.

1) Calculation of DSI: We set the region of interest (ROI) on the left (reference) image whose upper boundary is the vanishing line of the road plane and divide the image in the ROI into vertical columns of pixels (see Fig.8). Under the road environment constraint, a single disparity parameter determines the whole correspondence for a column of pixels in the reference image since the y-coordinate of the road-obstacle boundary can be calculated from (2).

The DSI is calculated by repeating the following procedure for every i and j:

1) Calculate the y-coordinate of the hypothetical road-obstacle boundary from the disparity value j and (2).
2) Calculate the matching score $\mathrm{dsi}_r(i, j)$ for the road region (pixels below the boundary) of the ith column with the correspondence given by (1).
3) Calculate the matching score $\mathrm{dsi}_o(i, j)$ for the obstacle region (pixels above the boundary) of the i th column where the correspondence is given by horizontal translation of j pixels.

The matching score of the road regions and that of the obstacle regions are stored separately for the convenience of the proceeding optimization procedure.

2) Matching Score: In order to calculate the DSI, we have to define the matching score which evaluate the goodness of the match. Most of the conventional stereo methods use matching score or cost based on the intensity difference between corresponding pixels or regions, such as sum of squared differences (SSD) and sum of absolute difference (SAD). However, intensity-based matching measures are not robust enough under practical conditions. SAD and SSD rely on the constant luminance assumption. Therefore, they are sensitive to differences in camera gain or bias. Normalized correlation can compensate bias and multiplicative variation but is sensitive to outliers. Besides, image sampling tends to cause large intensity differences in textured regions unless image registration is done with sub-pixel accuracy[2].

Compared to image intensity, the direction of an edge is stable under various lighting conditions since it is invariant with respect to bias and multiplicative variation. Since calculating gradients has blurring effect, we do not need sub-pixel

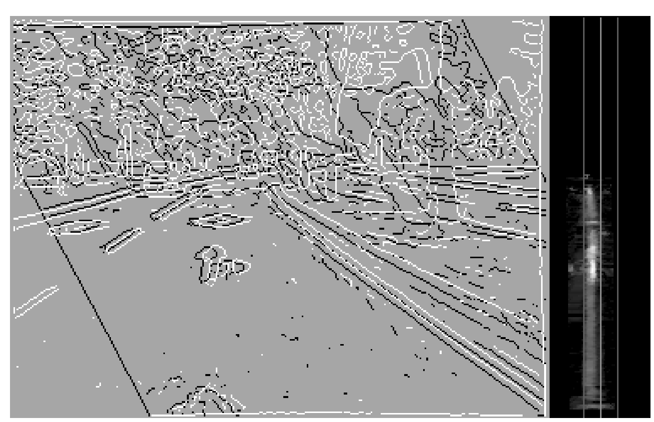

Fig. 6. Left: edges of a left image and the remapped right image. Right: edge correlation of each scan line with horizontal shift.

Fig. 7. Edges of a left image and the right image remapped with the estimated parameters

image registration for the matching score calculation.

We calculate match scores only at salient edge pixels using the following simple binary score that compares the gradient vectors of the corresponding pixels: if the angle between the two vectors is smaller than the predefined threshold then score is 1 and otherwise score is 0. Edges are detected with Canny edge detector though we do not use hysterisis thresholding. Since scores for the unmatched edges are 0, outliers are just ignored and do not significantly affect the result.

Since the road plane is not front-parallel to the image planes, the directions of the corresponding gradient vectors in the road regions do not match well[13]. Therefore, we prepare the affine transformed right image with (1) and use that image for comparison of the road region pixels.

By using this matching score, the criterion for the optimization becomes quite simple: to find the best path which gives the highest number of matched edge pixels.

C. Global Optimization Using Dynamic Programming

The DP calculation is done from the right most column to the left most column using the following recursive equations,

$$\begin{aligned} M_1(d_1) &= m_1(d_1), \\ M_i(d_i) &= m_i(d_i) \\ &+ \max_{d_{i-1}}\{M_{i-1}(d_{i-1}) - c_i(d_i, d_{i-1})\}, \end{aligned} \quad (3)$$

where $m_i(d_i)$ is the matching score for node (i, d_i) (i.e. the node that represent i th column with disparity value d_i) , $c_i(d_i, d_{i-1})$ is the cost of the path from node $(i-1, d_{i-1})$ to node (i, d_i), and $M_i(d_i)$ is the best score up to node (i, d_i).

Then, the best path is given as a sequence of disparity values $d_1^*, \cdots, d_W^*$ using the following recursive equations,

$$\begin{aligned} d_W^* &= \arg\max_{d_W} M_W(d_W), \\ d_i^* &= \arg\max_{d_i}\{M_i(d_i) - c_{i+1}(d_{i+1}^*, d_i)\}. \end{aligned} \quad (4)$$

Any obstacle area in the left image has a half-occluded area on its left and points inside the half-occluded area are invisible in the right image (see area A and A' in Fig.9). This poses the following two requirements on the optimization procedure.

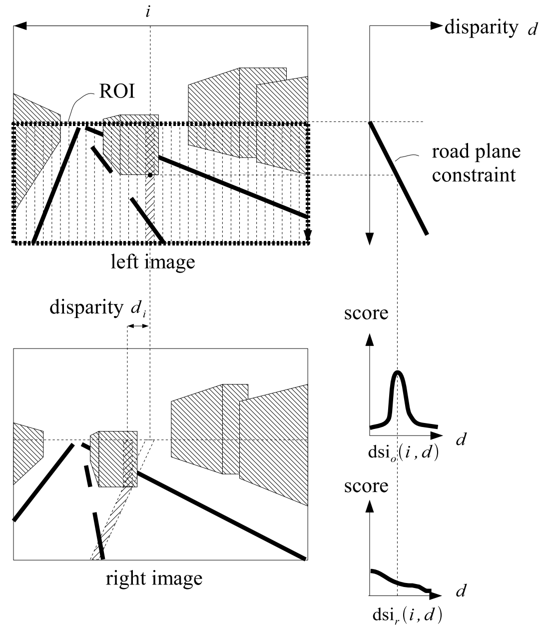

Fig. 8. Matching score calculation

- The road-obstacle boundary must not go inside half-occluded areas (which can be interpreted as the ordering constraint[3]), since the road-obstacle boundary should be visible in the both images.
- Half-occluded areas have to be excluded from the matching score calculations, since they do not have corresponding areas.

The above requirements restrict paths on DSI: no partial path can have slope any steeper than 45 degree toward upward-left and the partial path whose slope is 45 degree is considered to be a boundary between road area and half-occluded area. This constraint can be realized by the following matching score $m_i(d_i)$ and path cost $c_i(d_i, d_{i-1})$:

$$\begin{aligned} m_i(d_i) &= \mathrm{dsi}_r(i, d_i) + \mathrm{dsi}_o(i, d_i) \\ c_i(d_i, d_{i-1}) &= \begin{cases} \infty & \mathit{for}\ d_i < d_{i-1} - 1 \\ \mathrm{dsi}_o(i, d_i) & \mathit{for}\ d_i = d_{i-1} - 1 \\ 0 & \mathit{for}\ d_i > d_{i-1} - 1 \end{cases} \end{aligned} \quad (5)$$

The optimum path on the DSI is converted to the road-obstacle boundary in the left image using (2). The partial paths whose disparity values are smaller than a threshold or those which are considered to be boundary between road area and half-occluded areas are eliminated from the result. Fig.10 shows an example of a DSI and the estimated road-obstacle boundary.

D. Multiresolution Scheme

Our road environment model approximate obstacles with column-like structures as shown in Fig.4. In close distance, a small difference in depth causes a large difference in

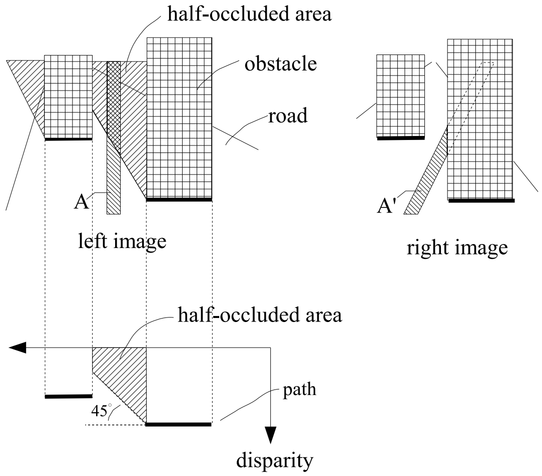

Fig. 9. Occlusions and the ordering constraint

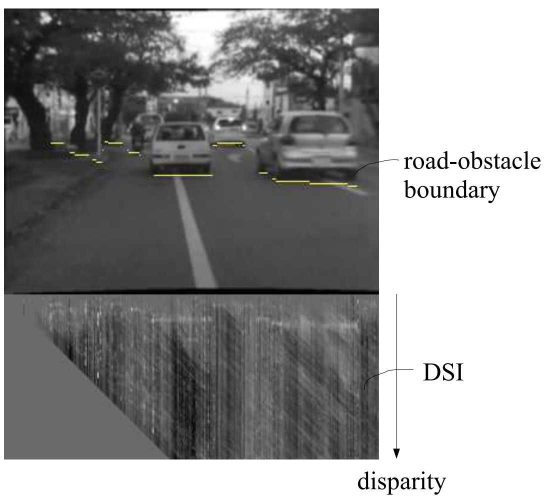

Fig. 10. DSI and road-obstacle boundary

disparity, which makes the approximation unfit. Therefore, close obstacles should be handled with coarse resolution while distant obstacles be handled with fine resolution. We use the multiresolurion scheme which adaptively chooses resolution for each column of the image.

Let l be the superscript that represent the scale with which images are reduced to $1/2^{l-1}$ of the original size ($l = \{1, \cdots, L\}$), dsi_r^l and dsi_o^l be the DSIs of scale l, and d_i^{*l} be the estimated disparity of the i th column of scale l.

We begin with the coarsest scale L and estimate the disparities d_i^{*L} for $i = \{1, \cdots, W/2^{L-1}\}$. In the subsequent scales, the disparity range for the i th column is limited to an interval $[2d_{i/2}^{*l+1} - \delta, 2d_{i/2}^{*l+1} + \delta]$, where $d_{i/2}^{*l+1}$ is the estimated disparity of the corresponding position in the preceding scale $l+1$ and δ is a small integer. If $d_{i/2}^{*l+1}$ is greater than threshold d_{th}, then, instead of calculating matching scores with current scale images, the values of DSIs in the preceding scale are used as follows:

$$\mathrm{dsi}^l(i, d) \quad = \quad 2\mathrm{dsi}^{l+1}(i/2, d/2). \tag{6}$$

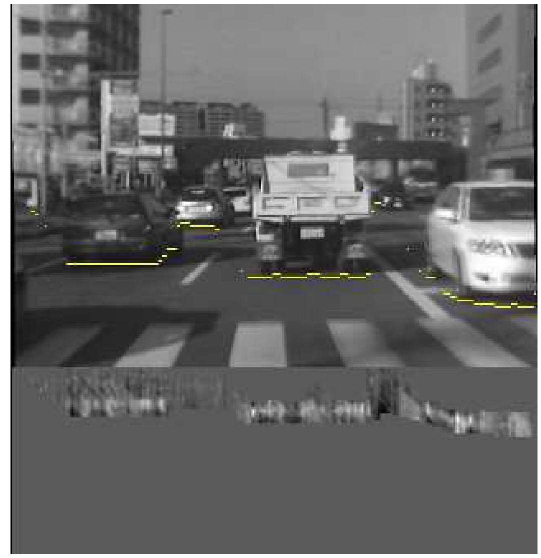

Fig. 11. An example of multiresolution processing

Fig.11 shows an example of a multiresolution processing, where the results of coarser levels adaptively limit the search range in disparity for each horizontal position.

IV. Experimental Results

We carried out experiments with various image sequences.

The proposed DP-based opitmization allows paths to have depth discontinuities while the ordering constraint prevent false-detection in half-occluded areas (see Fig.12).

Since the criterion for the optimization is quite simple, i.e. to find the road-obstacle boundary that maximize the number of matched edge pixels, the system works under various conditions without changing parameters. Fig.13 shows examples of night scenes.

Since the proposed matching measure is insensitive to bias and multiplicative variation, the system works stably under severe lighting conditions, such as shown in Fig.14.

Wet road surfaces cause reflections and the mirror images of obstacles appear to be negative obstacles which have negative height with respect to the road surface. The mirror images of obstacles and patterns on road surfaces, such as lane markings, cause multiple disparities in road areas, which pose difficulties for many conventional stereo methods. Since the proposed method is constrained to detect obstacles above the road surface, negative obstacles do not support any of the road-obstacle boundary hypotheses. Because of the robust matching measure which only counts the number of matched edge pixels, edges of mirror images of obstacles are ignored and do not significantly affect the results (see Fig.15).

Besides, the proposed method shows high tolerance to partial occlusions, such as ones caused by the wiper on the windshield (see Fig.16). The road environment constraint significantly reduces the degree of freedom of the disparity map so that neighboring areas can complement the lack of information caused by partial occlusions.

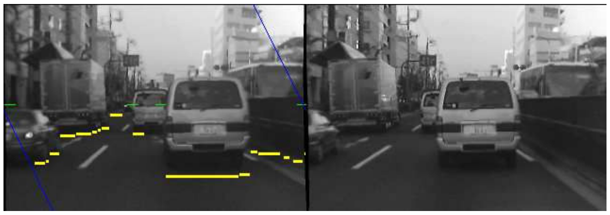

Fig. 12. An example of occlusions and depth discontinuities

Fig. 13. Examples of night scenes

Fig. 15. Examples of heavy rain

Fig. 16. An example of partial occlusion by wiper on the windshield

V. Conclusions

A fast and robust stereo algorithm for on-board obstacle detection systems has been proposed. The proposed method finds the optimum road-obstacle boundary under the road environment constraint. The road environment constraint along with the road plane constraint enables efficient DP-based global optimization.

Since the proposed method calculate matching score only at salient edge pixels and DP-based optimization does not require any iterations, the computational cost is small. Multiresolution scheme enables to handle a wide range of disparity. The processing time for VGA size image pair with 140 disparity range is about 15msec on a 3.6GHz pentium IV processor, which is fast enough for real-time applications.

References

[1] N. Ayache and C. Hansen, "Rectification of Images for Binocular and Trinocular Stereovision," *Proc. ICPR*, vol. I, pp.11-16, 1988.

[2] S. Birchfield and Carlo Tomasi, "Depth Discontinuities by Pixel-to-pixel Stereo," *Proc. ICCV*, pp.1073-1080, 1998.

[3] A. F. Bobick and S. S. Intille, "Large occlusion stereo," *International Journal of Computer Vision*, 33(3), pp. 181-200, 1999.

[4] M. Bertozzi and A. Broggi, "GOLD: A parallel real-time stereo vision system for generic obstacle and lane detection,"*IEEE Trans. Image Processing,* vol. 7, no. 1, 1998.

[5] H. Hattori and A. Maki, "Stereo wightout depth search and metric calibration," *Proc. CVPR*, vol. I, pp.177-184, 2000.

[6] H. Hattori and Nobuyuki Takeda, "Dense stereo matching in restricted disparity space," *IEEE Intelligent Vehicles Symposium,* pp.117-122, June 2005.

[7] R. Labayrade, D. Aubert, J. P. Tarel, "Real Time Obstacle Detection on Non Flat Road Geometry through V-Disparity Representation," it IEEE Intelligent Vehicules Symposium, pp.646-651, 2002.

[8] K. Y. Lee, J. W. Lee, and M. R. Cho, "Detection of road obstacles using dynamic programming for remapped stereo images to a top-view," *IEEE Intelligent Vehicles Symposium,* pp.765-770, 2005.

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

[10] Z. Sun, G. Bebis, and R. Miller, "On-Road Vehicle Detection: A Review," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 28, No. 5, pp. 694-711, 2006.

[11] H. Sunyoto, W. van der Mark, and D. M. Gavrila, "A Comparative Study of Fast Dense Stereo Vision Algorithms," *IEEE Intelligent Vehicles Symposium,* pp. 319-324, 2004.

[12] R. Y. Tsai, "An efficient and accurate camera calibration technique for 3D machine vision," *Proc. CVPR,* pp.364-374, 1986.

[13] T. Williamson and C. Thorpe, "Detection of Small Obstacles at Long Range Using Multibaseline Stereo," *IEEE Intelligent Vehicles Symposium*, pp. 311-316, 1998.

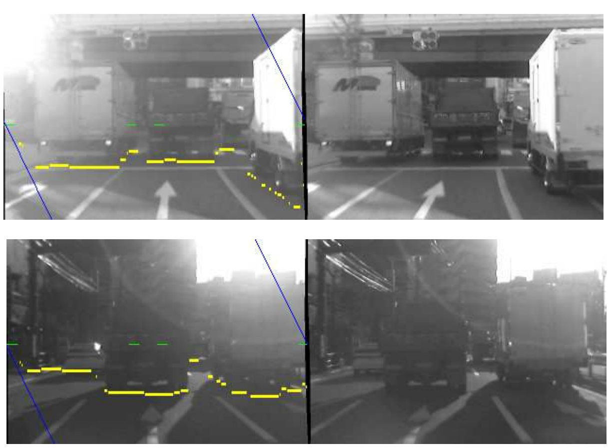

Fig. 14. Examples of severe lighting conditions