Road Scene Analysis by Stereovision: a Robust and Quasi-Dense Approach

Nicolas Hautière*, Raphaël Labayrade†, Mathias Perrollaz† and Didier Aubert†
*DESE - LCPC, 58 Boulevard Lefebvre, 75732 Paris Cedex 15, France
Email: hautiere@lcpc.fr
†LIVIC - INRETS / LCPC, Bldg 824, 14 route de la Minière, 78000 Versailles, France
Email: labayrade@lcpc.fr, {perrollaz, aubert}@inrets.fr

***Abstract*— A stereovision method is presented in this paper, to compute reliable and quasi-dense disparity maps of road scenes using in-vehicle cameras. It combines the advantages of the "v-disparity" approach and a quasi-dense matching algorithm. In this aim, road surface and vertical planes of the scene are first extracted using the sparse "v-disparity" approach. The knowledge of these global surfaces of the scene is then used to guide a quasi-dense matching algorithm and to propagate disparity information on horizontal edges. Both algorithms are presented and compared. Then, our approach is presented and examples of quasi-dense disparity maps are given. Finally, the efficiency of the method is illustrated by the accurate positioning of a bounding box around a vehicle in a bad contrasted video sequence.**

***Keywords*—ITS, stereovision, u-v disparity, quasi-dense matching, bounding box**

I. Introduction

Stereovision techniques aiming to analyse outdoor scenes are numerous. Among them, some techniques are devoted to in-vehicle obstacle detection [1]–[5]. Due to real-time constraints, most techniques rely on sparse matching techniques to estimate depth maps of the scene. This type of techniques poorly reconstruct the scene. Consequently, using a sparse disparity map, a direct segmentation of the scene is hazardous. To ensure a high rate of robustness, a solution is to use voting techniques which have already shown good properties. That's why "v-disparity" approach [4] which enables to compute the longitudinal road profile and to estimate the distance to objects above the road surface has met a certain success and is now widely used [6]–[8].

However, a generic and robust method based on stereovision, which provides an accurate lateral position of obstacles is still missing. Indeed, using rectified images, sparse techniques relying on horizontal gradients to match the stereo pairs can not entirely reconstruct objects composed of both vertical and horizontal gradients. Consequently, even using a robust technique like a "u-disparity" approach, left and right sides of obstacles are often disconnected leading to incomplete objects segmentations.

A LIDAR is often used to fill this gap. Thus, combined with a LIDAR which provides the lateral position of objects, our system based on the "v-disparity" approach has proved to be very efficient to detect near obstacles [9]. A collision mitigation system has even been demonstrated with a great success [10].

On the other hand, in the field of 3D reconstruction, some techniques are devoted to the computation of dense disparity maps [11]. However most of these techniques are very costly to implement.

Lhuilier [12] developed a quasi-dense matching algorithm, which can be considered as a in between method. Its principle is first to compute a sparse disparity map and then to perform a region growing in the disparity space to progressively densify the disparity map.

In this paper, we propose a method which allows to compute a quasi-dense and reliable disparity map, combining the "v-disparity" semi-global approach and the quasi-dense matching algorithm. In the following, both methods are successively described. Then, the proposed approach is presented and is applied to the accurate estimate of the position of obstacles. The techniques are illustrated by means of a bad contrasted road sequence (foggy weather), because such weather conditions are interesting to test the robustness of the different methods.

II. The "v-disparity" Approach

A. The Image of a Plane in the "v-disparity" Image

The stereovision algorithm uses the "v-disparity" transform, in which the detection of straight lines is equivalent to the detection of planes in the scene. In this aim, we represent the v coordinate of a pixel towards the disparity Δ (performing

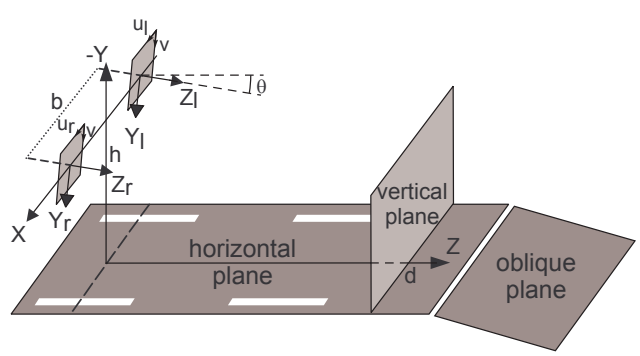

Fig. 1. Domain of validity of the study and coordinate systems used.

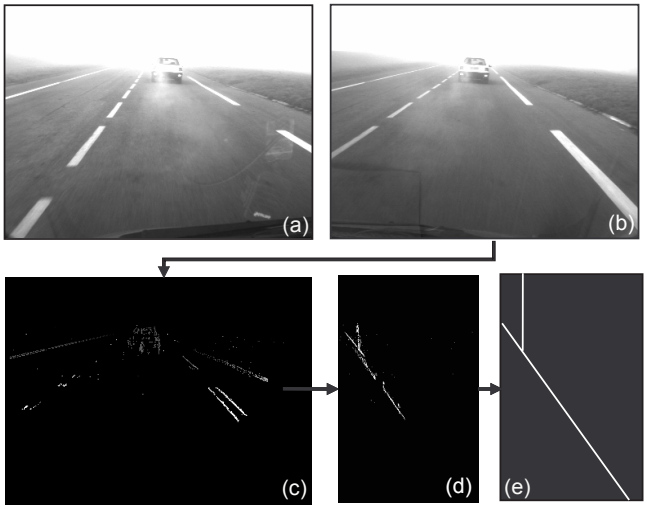

Fig. 2. Overview of the "v-disparity" framework using a stereo pair under foggy weather. (a) Left original image; (b) right original image; (c) rough disparity map computed from images (a) and (b); (d) "v-disparity" image; (e) extracted lines from the "v-disparity" image.

accumulation from the disparity map along scanning lines) and detect straight lines and curves in this "v-disparity" image (denoted by I_{v_Δ}) [4].

This algorithm assumes the road scene is composed of set of planes: obstacles are modelized as vertical planes, whereas the road is supposed to be an horizontal plane (when it is planar), or a set of oblique planes (when it is not planar), as shown in Fig. 1.

According to the modeling of the stereo sensor given in Fig. 1, the plane of equation $Z = d$, corresponding to a vertical object, is projected along the straight line of Eq. (1) in I_{v_Δ}:

$$\Delta = \frac{b}{d}(v - v_0)\sin\theta + \frac{b}{d}\alpha\cos\theta \tag{1}$$

The plane of equation $Y = 0$, corresponding to the road surface, is projected along the straight line of Eq. (2) in I_{v_Δ}:

$$\Delta = \frac{b}{h}(v - v_0)\cos\theta + \frac{b}{h}\alpha\sin\theta \tag{2}$$

The different parameters are as follows: (u, v) denotes the position of a point in the image, (u_0, v_0) is the projection of the optical center in the image, α is the ratio between the focal length and the size of pixels, θ is the angle between the optical axis of the cameras and the horizontal, h is the height of the cameras above the ground and b is the distance between the cameras (i.e. the stereoscopic base). Mathematical details can be found in [4].

B. "V-disparity" Image Construction and 3D Surface Extraction

The algorithm performs a robust extraction of these planes from which it deduces many useful information about the road and the obstacles located on its surface. Fig. 2 illustrates the outline of the process. From two stereo images (a) and (b), a disparity map I_Δ (c) is computed (Sum of Square Differences -SSD- criteria is used for this purpose along edges). The disparity values are represented by a grey level. Then an accumulative projection of this disparity map is performed to build the "v-disparity" image $I_{v\Delta}$ (d). For the image line i, the abscissa u_M of a point M in $I_{v\Delta}$ corresponds to the disparity Δ_M and its grey level i_M to the number of points with the same disparity Δ_M on the line i : $i_M = \sum_{P \in I_\Delta} \delta_{v_P,i}\delta_{\Delta_P,\Delta_M}$ where $\delta_{i,j}$ denotes the Kronecker delta.

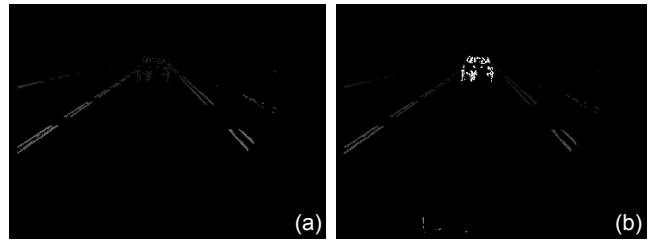

Fig. 3. (a) Improved disparity map (b) Results of obstacles areas detection.

From this "v-disparity" image, a robust extraction of straight lines is performed through a Hough transform. This extraction of straight lines (e) is equivalent to the extraction of the planes of interest taken into account in the modelization of the road scene.

C. Disparity Map Improvement

In order to quickly compute the "v-disparity" image, a sparse and rough disparity map has been built. This disparity map may contain numerous false matches, which prevent us to use it as a depth map of the environment. Thanks to the global surfaces extracted from the "v-disparity" image, false matches can be partially removed. In this aim, we check wether a pixel of the disparity map belongs to any global surface extracted using the same matching process. If it is the case, the same disparity value is mapped to the pixel and leads to Fig. 3a. Obstacles areas, that is to say objects above the road surface, can be deduced using the same matching process. This leads to Fig. 3b. Details of this process can be found in [13].

D. Lateral Position of Obstacles

In-vehicle stereovision approaches usually use a rectified geometry, in order to minimize the computational cost. Indeed, using such a configuration, matches are on same line of the stereo pair. Hence, sparse techniques rely only on horizontal gradients to match the stereo pairs. The disadvantage of this sensor configuration is that disparity can not be computed on horizontal edges. In this way, we can not entirely reconstruct objects composed of both vertical and horizontal gradients, like the backside of a vehicle for example.

To solve the problem, some solutions have been proposed using symmetry [14]. However, such a method is not generic. Labayrade [15] proposed to perform an accumulative projection of this disparity map along the horizontal axis, in order to build a "u-disparity" image $I_{u\Delta}$. Unfortunately, even using

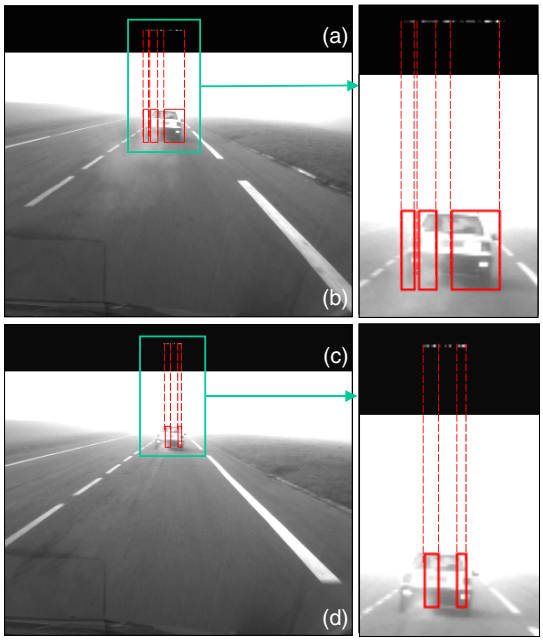

Fig. 4. (a)(c) Results of "u-disparity" image $I_{u\Delta}$ computations using only vertical edges; (b)(d)left and right sides of the vehicle are disconnected and lead to incomplete bounding boxes.

a robust technique like "u-disparity", left and right sides of obstacles are often disconnected leading to incomplete object segmentations.

This problem is illustrated in Fig. 4. In this figure, we present two images of a test video sequence and the corresponding "u-disparity" images. The horizontal segments, which correspond to the projection of the vehicle have some holes. Hence, if we detect the different segments in order to position a bounding box around the vehicle, this lead to incomplete segmentations, which is problematic.

In the following, we are going to propose a solution based on a quasi-dense matching algorithm to densify the disparity map and solve this problem.

III. The Quasi-Dense Matching Approach

Lhuilier [12] developed a quasi-dense matching algorithm. Its principle is first to compute a sparse disparity map and then to perform a region growing in the disparity space to progressively densify the disparity map.

A. Seed selection

The first step of the algorithm is the computation of a sparse disparity map using the $ZNCC$ correlation measure:

$$\frac{\sum_i \left(I(x+i)-\bar{I}(x)\right)\left(I'(x+\Delta+i)-\bar{I}'(x+\Delta)\right)}{\sqrt{\sum_i \left(I(x+i)-\bar{I}(x)\right)^2 \sum_i \left(I'(x+\Delta+i)-\bar{I}'(x+\Delta)\right)^2}} \tag{3}$$

where $\bar{I}(x)$ and $\bar{I}'(x+\Delta)$ are the means of pixel intensities for the window centered at x and Δ is the considered shift.

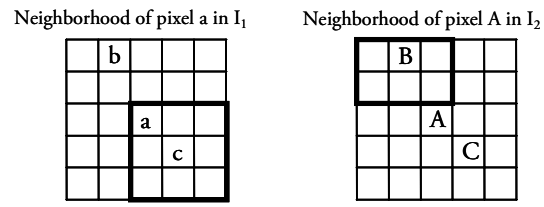

Fig. 5. Definition of neighborhood $N(a, A)$ of pixel match (a, A). It is a set of matches included in the two 5×5-neighborhood $N_5(a)$ and $N_5(A)$ of pixels a and A. Possible matches for b (resp. C) are in the 3×3 black frame centered at B (resp. c).

A cross correlation is then used to reduce the number of false matchings.

B. Propagation

The idea consists in propagating the initial seeds in a way similar to a region growing, guided not by a criterion of homogeneity but by a score of correlation. All seed matches are the starting point of concurrent propagations. At each stage, a match (a, A) with the best $ZNCC$ score is removed from the current set of seed matches. Then new matches are searched in the "match neighborhood" and are added to the current set of seeds and to the set of accepted matches. The neighbors of pixels a and A are taken to be all pixels within the 5×5 window centered at a and A (cf. Fig. 5). For each neighboring pixel in the first image, the possible match candidates are all pixels of a 3×3 window in the neighborhood of its corresponding location in the second image. One thus incites the gradient of displacements not to exceed a pixel.

Lhuillier also defines $s(x) = \max\{|I(x+\Delta) - I(x)|, \Delta \in N_4(x)\}$, an estimate of the luminance roughness for the pixel at x, which is used to stop propagation into insufficiently textured areas a with $s(a) < t$ and $0 \leq I(a) \leq 1$. $I(a)$ denotes the normalized intensity value for the pixel at a of image I. A typical value for t is 0.01.

C. Application to Road Scene Analysis

Compared to classical stereo images, typical road images are textureless, especially when the road scene is bad con-

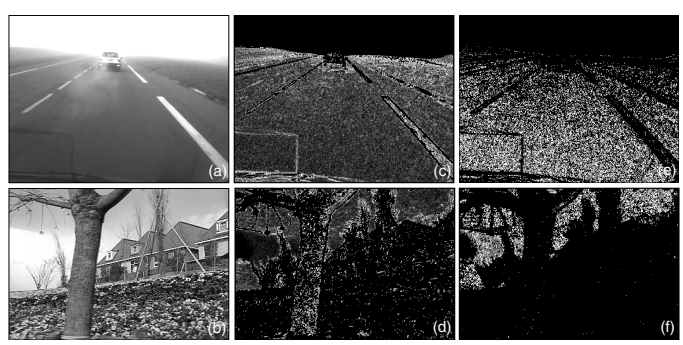

Fig. 6. Comparison of roughness of (a) our road image under foggy weather; (b) image used by Lhuillier; (c)(d) white pixels x where $0 < s(x) < 0.05$; (e)(f) white pixels x where $0 < s(x) < 0.01$.

trasted like under foggy weather. To prove it, we compared the roughness of two images. Fig. 6a is our bad contrasted road image under foggy weather. Fig. 6b is a classical image in computer vision and is used by Lhuillier [12] to illustrate his algorithm. The areas, where pixels are low textured ($0 < t < 0.05$) and very low textured ($0 < t < 0.01$), are respectively given in white in Figs. 6cd and Figs. 6ef. We can see in it that the road surface is very low textured compared to the classical image. Consequently, the propagation will be quickly stopped on low textured areas and it will be difficult to recover disparity information on the entire road surface.

We have applied the quasi-dense matching algorithm to our bad contrasted stereo pair. The idea is first to compute a quasi-dense disparity map and thereafter to compute the "v-disparity" image, so as detect the road surface and the obstacles.

Thus, Fig. 7 gives different results of disparity map computations using different values of t. The corresponding "v-disparity" images are also given. We can state that if the threshold t is too low, too much correlated matching errors occur and prevent us to accurately detect the global surface of the image using the "v-disparity" approach.

Indeed, in Fig. 7d, the straight lines corresponding to the road shape and the vertical plane of the vehicle can be detected, because correlated matching errors are not too numerous in Fig. 7a. On the contrary, in Fig. 7e and Fig. 7f, the straight lines corresponding to the road surface and the vertical segment are too thick to be accurately detected. This is due to the fact that the road is textureless. Consequently, $ZNCC$ criteria is not enough discriminant in order to ensure a good matching.

Finally, whereas the quasi-dense matching algorithm enables to propagate disparity along horizontal edges, this experiment shows that it is not adapted to compute a reliable disparity map of a road scene.

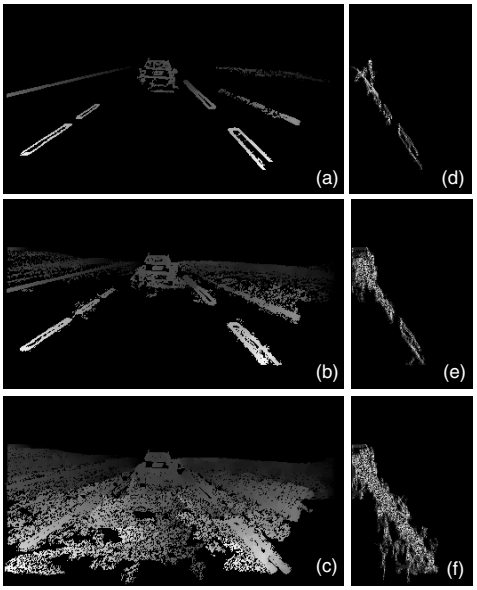

Fig. 7. Results of the quasi-dense matching algorithm on our test images for different values of t: (a) $t = 0.05$, (b) $t = 0.02$, (c) $t = 0.01$. (d)(e)(f) corresponding "v-disparity" images. When the threshold t is lowered, numerous correlated errors appear and straight lines in the "v-disparity" image get thicker.

IV. Robust and Quasi-Dense Approach

A. Problem Statement

On one hand, the "v-disparity" approach is quite robust, thanks to the accumulation technique used. So far, the method, which rely only on horizontal gradients to match the stereo pairs, can not compute the disparity on the horizontal edges. On the other hand, the quasi-dense matching algorithm allows to propagate disparity information along both vertical and horizontal edges. However, it is difficult to obtain a dense and accurate disparity information on the road surface, because it is textureless (Fig. 6). The threshold t which stops the propagation can be lowered. Unfortunately, this leads to correlated errors. Thus, the road shape in the "v-disparity" image is thicker, like the vertical objects. Road shape and obstacles can not be accurately detected (cf. Fig. 7ef).

B. Computation of a Reliable and Quasi-Dense Disparity Map

To compute a better disparity map, we then propose to use the advantages of both previously described methods. We use the "v-disparity" approach to estimate the longitudinal profile of the road and to detect the presence of obstacles. The quasi-dense matching algorithm is used to propagate the disparity information along the horizontal edges.

Thus, seed matches of the quasi-dense matching algorithm are used to compute the "v-disparity" image, in order to extract the road surface and the positions of the different vertical planes of the scene. Then, we propagate the initial seeds like Lhuillier does [12], except that for each match candidate we check if it belongs to one of the planes of the "v-disparity" image. If it is the case, the match candidate is added to the current set of seeds and to the current set of accepted matches, which is under construction. Otherwise, the match candidate is removed from the current set of seeds.

Compared to the previous section, the propagation is done after the computation of the "v-disparity" image. Because the computation of this image rely on the matches of vertical edges, its computation is robust as well.

In this way, we add a global constraint to the quasi-dense matching algorithm. Thus, the number of correlated errors is drastically reduced and the number of pixels belonging to a global surface with known disparity is increased. Disparity information can be propagated along the horizontal edges without changing the estimated geometry of the scene.

C. Results

This method is illustrated in Fig. 8 for different values of the threshold t. Reconstructed "v-disparity" images are given. Contrary to Fig. 7, the number of matching errors is reduced and dense disparity information is obtained on the road surface. Reconstructed "v-disparity" images are good and do not differ if the threshold t is lowered. However, as previously said,

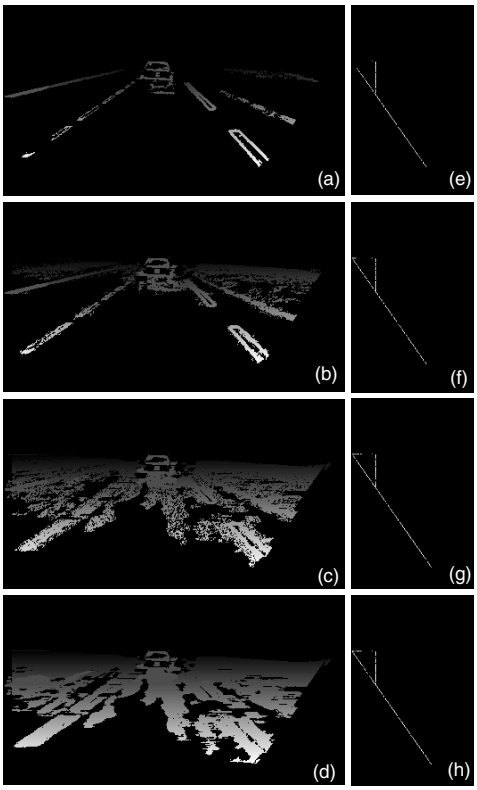

Fig. 8. Results of our method on our test images for different values of t: (a) $t = 0.05$, (b) $t = 0.02$, (c) $t = 0.01$, (d) $t = 0.001$. (e)(f)(g)(h) Corresponding "v-disparity" images.

disparity information in front of the vehicle is difficult to obtain because the image is very low textured (cf. Fig. 6f). There are remaining matching errors due to half occluded contours. However, numerous methods are devoted to the detection of such contours [16].

V. Application to the Lateral Position of Obstacles

Once we have a quasi-dense disparity map, we can use it to detect the lateral position of obstacles and compare the results with those presented in section II-D.

A. Robust "u-disparity" Image Computation

Thanks to the proposed method, the disparity map is quasi-dense, especially on the vertical objects, like on the vehicle of our test image. We can now compute a quasi-dense and reliable "u-disparity" image. A sample is given in Fig.9b. Compared with Fig. 4, the obtained "u-disparity" image is much better, because vertical and horizontal edges are taken into account to compute it. Consequently, the lateral segmentation of the object is easier.

B. Computation of a Bounding Box around Objects

To position an accurate bounding box around an object, we need three parameters: the longitudinal position, the height and the lateral position of this object. The longitudinal position of objects is given by the vertical plane in the "v-disparity" image. The height is also computed thanks to the "v-disparity" image. Details can be found in [15].

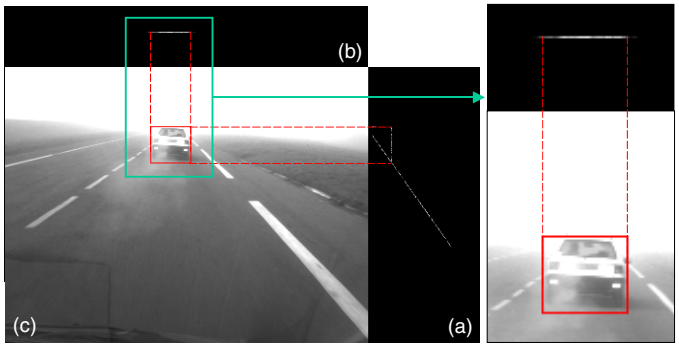

Fig. 9. Result of (a) "v-disparity" image (b) "u-disparity" image computations using our modified quasi-dense matching algorithm. (c) This leads to the computation of robust and precise bounding boxes around obstacles without any arbitrary thresholds.

The lateral position is given by the "u-disparity" image. Indeed, it is enough to detect the horizontal segment, which is now quasi-dense, to ensure a good detection of the lateral position of the vertical objects. Finally a good positioning of the bounding boxes can be made, like in Fig. 9. Thus, using this approach, the risk that the same object is split into different bounding boxes is reduced.

C. Results

Our method has been tested using a bad contrasted video sequence, where a vehicle is moving away. In Fig. 10, we give some samples of this video sequence. Although the vehicle is bad contrasted, the bounding box is correctly positioned, even when the vehicle is quite far away.

We have also applied our technique to other types of scenes. In Figs. 11ab, we present some results with a pedestrian and a vehicle which illustrates the genericity of the approach. In Figs. 11cd, we present some results with two cars on a paved road, where classical disparity maps usually contain a lot of false matches because of the periodic texture of the road.

VI. Discussion and Future Work

The proposed approach enables to solve the problem of disparity computation along horizontal edges, which is a crucial problem for in-vehicle stereovision techniques. Hence, it allows us to position more precisely bounding boxes around different types of road objects. There are three mainly remaining problems to solve. Firstly, the choice of the threshold t is currently arbitrary chosen. The computing time is directly related to this value, which is a crucial point for such a real-time application. Secondly, two close objects which belongs to the same "v-disparity" plane can be aggregated within the same bounding box, because their respective segments in the "u-disparity" image can be confused, in particular in case of distant objects. Thirdly, a strategy to mitigate the influence of half-occluded contours in the computation of "u-disparity"

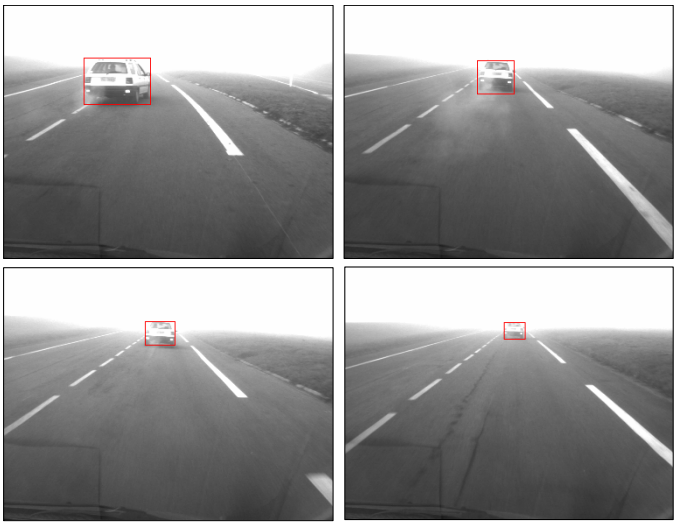

Fig. 10. Examples of bounding boxes around a vehicle in foggy weather at various distances. Although the vehicle is hardly visible and can be confused with the sky, accurate bounding boxes are provided.

images must be implemented. In the future, it is important to solve the aforementioned problems. Then, the next step is a quantitative assessment of the method compared with other state-of-the-art schemes.

VII. Conclusion

In this paper, we present a stereovision based technique, which enables to compute a quasi-dense and reliable disparity map of road scenes. The method has the advantages of the "v-disparity" approach and of a quasi-dense matching algorithm, which allows us to propagate disparity information along horizontal edges. The three different approaches are compared and the robustness of our approach is demonstrated through use of a bad contrasted video sequence under foggy weather. Finally, this technique allows us to compute a quasi-dense "u-disparity" image, which leads to a generic and robust method to detect the lateral position of vertical objects. In the future, we would like to use this disparity map in other applications, such as contrast restoration.

References

[1] J. Weber, D. Koller, Q.-T. Luong, and J. Malik, "An integrated stereo-based approach to automatic vehicle guidance," in *International Conference On Computer Vision*, 1995.

[2] T. A. Williamson, "A high-performance stereo vision system for obstacle detection," Ph.D. dissertation, Robotics Institute Carnegie Mellon University, Pittsburg, 1998.

[3] M. Bertozzi and A. Broggi, "GOLD: a parallel real-time stereo vision system for generic obstacle and lane detection," *IEEE Transactions on Image Processing*, vol. 7, no. 1, pp. 62–81, 1998.

[4] D. Aubert and R. Labayrade, "Road obstacles detection by stereovision: the "v-disparity" approach," *Annals of Telecommunications*, vol. 60, no. 11–12, 2005.

[5] P. Lombardi, M. Zanin, and S. Messelodi, "Unified stereovision for ground, road, and obstacle detection," in *IEEE Intelligent Vehicles Symposium*, 2005.

[6] A. Broggi, C. Caraffi, R. I. Fedriga, and G. P., "Obstacle detection with stereo vision for off-road vehicle navigation," in *Procs. Intl. IEEE Wks. on Machine Vision for Intelligent Vehicles*, 2005.

[7] H. Zhencheng and K. Uchimura, "U-v-disparity: An efficient algorithm for stereovion based scene analysis," in *IEEE Intelligent Vehicles Symposium*, 2005.

[8] L. Petersson, L. Fletcher, A. Zelinsky, N. Barnes, and F. Arnell, "Towards safer roads by integration of road scene monitoring and vehicle control," *International Journal of Robotics Research*, vol. 25, no. 1, pp. 53–72, 2006.

[9] R. Labayrade, C. Royere, D. Gruyer, and D. Aubert, "Cooperative fusion for multi-obstacles detection with use of stereovision and laser scanner," *Autonomous Robots*, vol. 19, no. 2, pp. 117–140, 2005.

[10] R. Labayrade, R. Royère, and D. Aubert, "A collision mitigation system using laser scanner and stereovision fusion and its assessment," in *IEEE Intelligent Vehicles Symposium*, 2005.

[11] W. Van der Mark and D. M. Gavrila, "Real-time dense stereo for intelligent vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 1, pp. 38–50, March 2006.

[12] M. Lhuillier and L. Quan, "Match propagation for image-based modeling and rendering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1140–1146, 2002.

[13] R. Labayrade and D. Aubert, "In-vehicle obstacles detection and characterization by stereovision," in *Proceedings of the 1st International Workshop on In-Vehicle Cognitive Computer Vision Systems, Graz, Austria*, November 2003.

[14] M. Bertozzi, A. Broggi, A. Fascioli, and S. Nichele, "Stereo vision-based vehicle detection," in *IEEE Intelligent Vehicles Symposium*, 2000.

[15] R. Labayrade, D. Aubert, and J.-P. Tarel, "Real time obstacle detection in stereovision on non flat road geometry through v-disparity representation," in *IEEE Intelligent Vehicles Symposium*, 2002.

[16] G. Egnal and R. P. Wildes, "Detecting binocular half-occlusions: Empirical comparisons of five approaches," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, 2002.

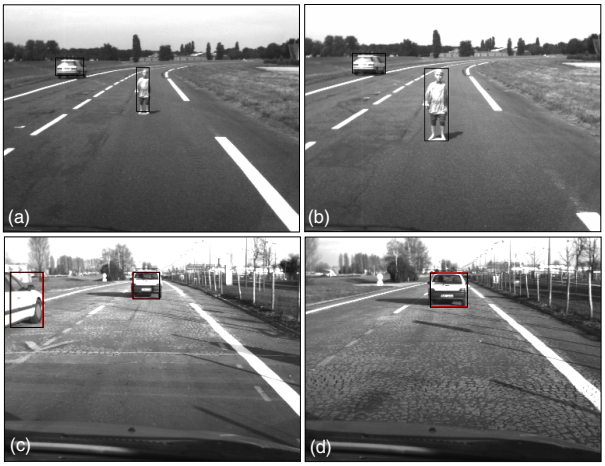

Fig. 11. Examples of bounding boxes around multiple objects: (a)(b) a car and a pedestrian; (c)(d) two cars on a paved road, which is a typical periodic texture.