HAL
open science

Weighted V-disparity Approach for Obstacles Localization in Highway Environments

Nizar Fakhfakh, Dominique Gruyer, Didier Aubert

To cite this version:

Nizar Fakhfakh, Dominique Gruyer, Didier Aubert. Weighted V-disparity Approach for Obstacles Localization in Highway Environments. IEEE Intelligent Vehicles Symposium, Jun 2013, Australia. 8p. hal-00865766

HAL Id: hal-00865766

https://hal.science/hal-00865766v1

Submitted on 25 Sep 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Weighted V-disparity Approach for Obstacles Localization in Highway Environments

Nizar Fakhfakh[1], Dominique Gruyer[1] and Didier Aubert[2]

***Abstract*— The employment of embedded passive sensors in order to perceive environment for reducing the accident risk level is a tendency of intelligent vehicles research. From such sensors, one can extract useful informations which can assist the driver to identify hazardous situations. While safety improvement is a substantial requirement for driving assistance, localizing and tracking obstacles in complex road environment became an important task. Stereovision is an attractive techniques which allows obtaining the 3D components of an observed scene from two visible 2D images. One promising approach is to use the V-disparity technique. It is a cumulative space estimated from the disparity image. We propose a sound framework and a complete system based on a real-time stereovision for detection, 3D localization and tracking of dynamic obstacles in highway environment. The main contribution we propose is the improvement of the V-disparity approach by extending the basic approach by merging it with a confidence term. This consists on weighting each pixel in the V-disparity space according to a confidence value which measures the probability of associating a pair of pixels. Furthermore, we propose a tracking system which is based on the belief theory. The tracking task is done on the image space which takes into account uncertainties, handles conflicts, and automatically dealt with targets appearance and disappearce as well as their spatial and temporal propogation. Extensive experiments on simulated and real dataset demonstrate the effectiveness and the robustness of the weighted V-disparity approach.**

I. INTRODUCTION

Establishing vehicle-infrastructure-driver interactions depends on the effectiveness of the perception task in a specific road scenario. The outstanding development of hardware have significantly improved the capabilities of embedded algorithms for accidents avoidance and drivers' safety improvement. One of the early researche works introducing stereovision for road obstacles detection and road-line markings extraction is the one that is proposed in [1]. Unfortunately, this previous work has proposed a sparse disparity map and the detection of obstacles is not performed in real time. The use of binocular stereopsis principle for collision avoidance or mitigation has been first successfully demonstrated by JPL's planetary robotic vehicle [2] which presents a trade-off between the computational time and the density of the disparity map. The authors in [3] uses vision-based detectors for estimating of the road profile. This is done by detecting road-line markings and all of obstacles not lying on the road plane are discarded. In [4] the authors have proposed an approach for the construction of occupancy grids using stereo camera pair for obstacles detection implemented on a GPU. Their method considers the geometrical visibility of the different regions of the image and can deal with partially-occluded objects. In [5] the regions so that a fixed-size volume contains a sufficient number of points are considered as potential obstacles.

In this paper, we propose a real-time video-based obstacles localization and tracking system at daytime which makes extensive use of stereovision principle. The proposed system allows detecting, localizing and tracking obstacles in highway environments in daytime. After a digitalization and a pre-processing stage, a stereo algorithm based on a local approach is then performed for dense disparity map generation from which the U- and V-disparity maps are built. The proposed work fits into earlier work initiated in [6] which exploits the cumulative spaces termed U- and V-disparity. The main contribution of this paper is in the improvement of the V-disparity approach by extending the basic approach by merging it with a confidence approach. This consists on weighting each pixel in the V-disparity space according to a confidence value which measures the probability of associating a pair of pixels. The two maps, U and V-disparity, are useful for estimating the 2D bounding box for each potential obstacle. A set of targets is therefore obtained with a state vector for each target. Each state vector contains the lateral and longitudinal position, the bounding box parameters and the uncertainty information. We have introduced a tracking filter with an association step based on belief theory and a filtering step using a linear Kalman filter. The association step allows to provide a link between a target and a track. The filter provide an update of the track state vector given the target data. This allows decreasing false positive rate and avoiding missed detections. The tracking task is done on the image space and allows taking into account uncertainties, handles conflicts, and automatically dealt with the appearance and the disappearing of targets as well as their spatial and temporal propogation. Then, the objects which are not lying in the road plane and hence non-potential obstacles are removed. The remaining potential obstacles are kept and are subject for further treatment.

Section II provides a brief overview of the proposed system. Section III gives the first contribution of this work with the presentation of an improved dense disparity map building with confidence approach applied to each pixel. This confidence approach is used to improve the V-disparity approach. Section IV explains the second contribution dedi-

[1]N. Fakhfakh and D. Gruyer are with French Institute of Science and Technology for Transport, Development and Networks, LIVIC unit. 14, route de la Minière, Bât 824. -78000 Versailles - Satory. {nizar.fakhfakh, dominique.gruyer}@ifsttar.fr

[2]D. Aubert is with French Institute of Science and Technology for Transport, Development and Networks, LEPSIS unit. 58, boulevard Lefebvre, 75732 Paris Cedex 15. didier.aubert@ifsttar.fr

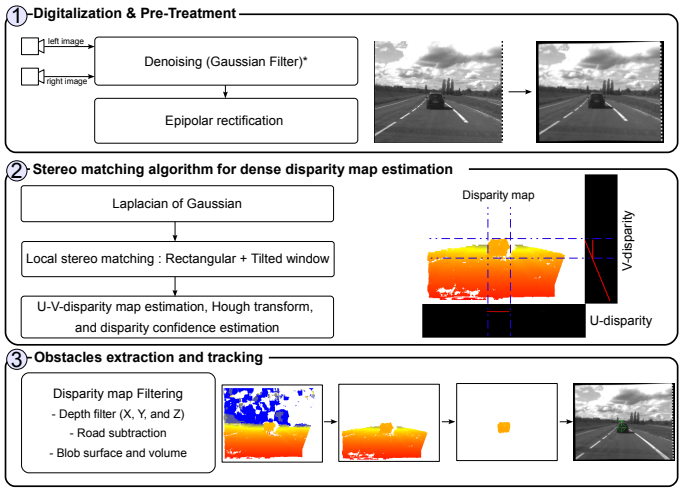


Fig. 1. Obstacles localization system.

cated to the tracking stage using probabilistic approach and the belief theory. Section V describes the implementation both in simulated and real scenarios, displays results, and evaluated the quality of this method with specifics scenarios related to the intelligent vehicle application. This evaluation part provides a third contribution with the use of automatic ground truth generation with the SiVIC virtual environment. Finally, section VI concludes and discusses future work.

II. PROPOSED OBSTACLES LOCALIZATION SYSTEM

Figure 1 illustrates the global block diagram of our obstacles localization system based on stereovision principle. The first step consists of the digitalization task and some pre-processing applied on both the left and the right images. The second and third sub blocks allows detecting and localizing obstacles by means of a robust stereo matching algorithm. All of the blocks are described in detail in the following sub sections.

A. Likelihood function for disparity map estimation

Estimating the real position of each point depends only on one unknown parameter which is the disparity measurement obtained usually by applying a stereo matching algorithm. The stereovision problem, also known as stereopsis problem, consists in determining the depth of all the objects in the image from the viewer, according to two images coming from the left and right cameras. The depth of any pixel is inversely proportional to its disparity, namely how much a pixel in the right image is shifted to obtain the corresponding pixel in the left image.

In order to reduce the complexity of such a stereo algorithm, one can exploit prior knowledge of the observed scene and cameras configuration, which allows promptly achieving an optimal solution to the matching task. Actually, known camera geometry with respect to the locally planar road is used to localize objects which are not lying in the road plane and hence can be considered as potential obstacles. An important step is the calibration process of the stereo cameras enables estimating the intrinsic and extrinsic parameters which will be useful for reducing the huge range of the disparity space search for each pixel. By means of a mechanical and soft calibration process, the stereo camera is then rectified in such a way that epipolar line appears horizontal in both images. Actually, the search for similarity between the two images can be reduced in this way to horizontal scanlines in the images. Currently, a correlation window is shifted for each pixel along a horizontal scanline and computes correlation measurement between the left and the right images. Each pair of pixels having an optimal score is considered and then a disparity measure is computed.

The local stereo matching we propose is a two steps patch matching process. For each pixel to be matched, two scores are obtained by applying once a rectangular window and once using a tilted window. This choice is motivated by considering the following observations:

– For a pixel which belongs to a vertical fronto-parallel plane, the score obtained with a rectangular window is less than the one obtained with a tilted window.

– For a pixel lying on the road plane, the score obtained with a tilted window is less than the one obtained with a rectangular window.

Local methods are known to rapidly achieve the stereo matching problem in reasonable processing time but are subject to many matching error. The estimation of disparities according to global approaches such as graphical model is known to be NP-hard, or computationally intractable problem. Another way for reducing the complexity of such a stereo algorithm is in exploiting prior knowledge of the observed scene and cameras configurations.

B. Confidence Approach for Disparity Characterization:

Each pair of matched pixels is evaluated by referring to the idea initiallty described in our recent and previous work [7]. Once a likelihood function is applied to initialize the disparity map, for each matched pair of pixels a confidence measure is computed. It is termed $\psi(p_l^{u,v}, p_r^{u',v})$ which represents the level of certainty considering a label l, i.e. disparity, as the best label for pixel p. We will start from an example given by figure 2 which shows a pixel which corresponds to a point of interest to be matched. The homologous of this pixel in the other image is unique and is well distinguished from all other candidate pixels. The proposed disparity confidence function gives a value scaled within the [0, 1] range. For the previous example, the proposed confidence function gives a real-valued number which is close to one. The closest to zero, the less confidence is it. This function depends on several local parameters and is given by Equation 1:

$$\psi(p_l^{u,v}, p_r^{u',v}) = P(p_r^{u',v}/p_l^{u,v}, \rho, min, \sigma, \omega) \qquad (1)$$

In order to well understand the parameters detailed later, figure 3 illustrates the scores obtained by applying the likelihood function on the pixel to be matched of figure 2. Each score obtained with a candidate qualifies the correlation degree with the pixel to be matched in the reference

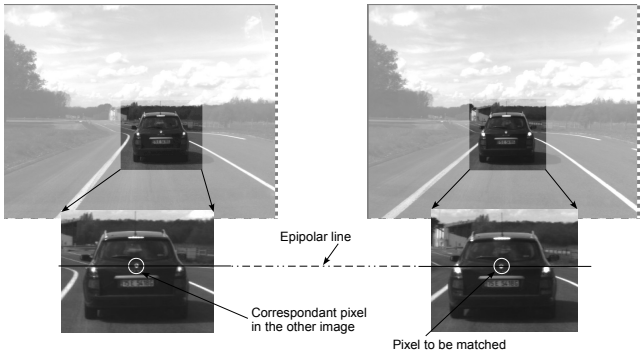
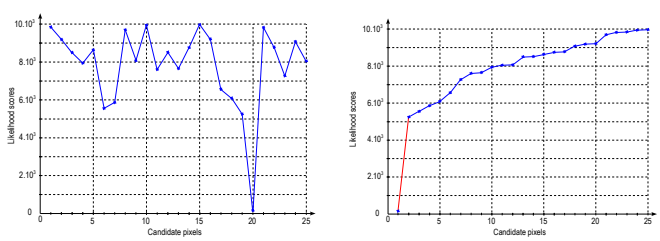


Fig. 2. An example of a point of interest to be matched from the right image which visibilly have one well identified correspondant in the left image.

image. The right-hand curve of the figure 3 represents the scores obtained with the likelihood function. Their scores are ordered and are shown in the left-hand curve of the figure 3.



(a) (b)

Fig. 3. Likelihood scores of the pixel of figure 2 obtained for each candidate pixels (a) non-ordered scores (b) ordered scores.

The confidence measurement with its parameters is given by Equation 2:

$$\psi(p_l^{u,v}, p_r^{u',v}) = \left(1 - \frac{min}{\omega}\right)^{\tau^2 log(\sigma)} \tag{2}$$

Where *min* is the Best Correlation Score, τ is the number of potential candidate pixels among all of the candidates, σ is the standard deviation of the disparity of the τ pixels, ω is the Gap between the τth and $(\tau+1)^{th}$ scores.

– *Best Correlation Score (min):* The output of the likelihood function is a measure of the degree of similarity between two pixels. Then, the candidate pixels are ranked in increasing order according to their corresponding scores. The couple of pixels that has the minimum score is considered as the best-matched pixels. The lower the score, the better the matching. The nearer the minimum score to zero, the greater the chance of the candidate pixel to be the actual correspondent.

– *Number of Potential Candidate Pixels (τ):* This parameter represents the number of potential candidate pixels having similar scores. τ has a big influence because it reflects the behavior of the likelihood function. A high value of τ means that the first candidate pixel is located either in a uniform region or in a repetitive pattern of the image. The lower the value of τ, the fewer the candidate pixels. If there are few candidates, the chosen candidate pixel has a greater chance of being the actual correspondent. Indeed, the pixel to be matched belongs to a region with high variation of gray component. A very small value of τ and a *min* score close to zero, means that the pixel to be matched probably belongs to a region of high gray variations.

– *Disparity variation of the τ pixels (σ):* A disparity value is obtained for each candidate pixel. For the τ potential candidate pixels, we compute the standard deviation σ of the τ disparity values. A small σ means that the τ candidate pixels are spatially neighbors. In this case, the true candidate pixel should belong to a particular region of the frame, such as an edge or a transition point. Therefore, it increases the confidence measure. A large σ means that the τ candidate pixels taken into account are situated in a uniform gray region.

– *Gap value (ω):* This parameter represents the difference between the τth and $(\tau+1)^{th}$ scores given with the dissimilarity function used. It is introduced to adjust the impact of the minimum score.

To ensure that function ψ has a value between 0 and 1, a few constraints are introduced. The *min* parameter have not to be higher than ω. If so, parameter ω is forced to $min+1$. Moreover, the $log(\sigma)$ term is used instead of σ, so as to reduce the impact of high value of σ and obtain coherent confidence measures.

In the following, the way of estimating the τ parameter is detailed. The number τ of potential candidate pixels is deduced from all scores obtained with the local likelihood function. The main idea is to detect major differences between successive scores. These major differences are called main gaps. Let ϕ denote a discrete function which represents all the scores given by the dissimilarity function in increasing order. We introduced a second function denoted η, which represents the average growth rate of the ϕ function. η can be seen as the ratio of the difference between a given score and the first score, and the difference between their ranks. This function is defined in Equation 3:

$$\eta(\phi^{u',v}) = \frac{\phi^{u',v}(z_m) - \phi^{u',v}(z_1)}{z_m - z_1} \qquad m \in \mathscr{L} \tag{3}$$

where $\mathscr{L}$ is the number of all candidates and $\phi^{u',v}(z_m)$ is the mth dissimilarity cost among all scores obtained for the pair of matched pixels $(p^{u,v}, p^{u',v})$. z_m is the rank of the mth score. $\eta(\phi^{u',v})$ is a discrete function that allows to highlight the large gaps between scores. It is materialized using Equation 4:

$$\xi(\phi^{u',v}) = \begin{cases} \frac{\nabla\eta^{u',v}}{m^2} & \text{if } \nabla\eta^{u',v} \geqslant 0 \\ -1 & \text{otherwise} \end{cases} \tag{4}$$

The previous function (Equation 4) is used to characterize the major scores and is applied only in the case where the gradient $\nabla\eta^{u',v}$ has a positive sign. We have introduced parameter m^2 in order to penalize the candidate pixels

according to their rank. The number of candidate pixels is given by Equation 5:

$$\tau = \arg\max_{m} \xi(\phi^{u',v}) \tag{5}$$

Figure 4 illustrates an example of estimating the number of potential candidate pixels τ by applying the ξ function on the scores showed in figure 3. In this example there is only one pixel which has a distinguished score among all scores. This is visibly highlighted in figure 4 so that the τ parameter corresponds to the rang of the higher value.

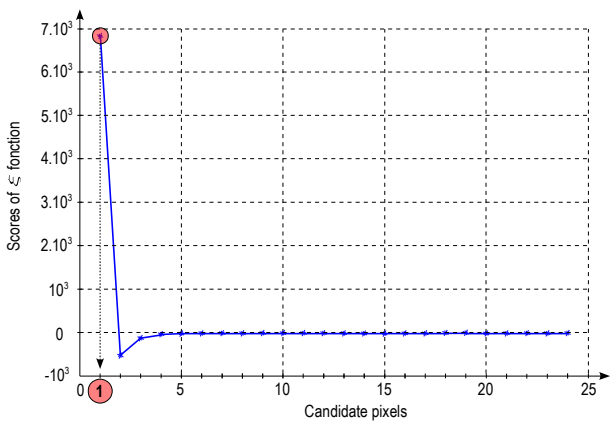


Fig. 4. The number of potential candidate pixels τ is given as the rank of the highest number obtained with the ξ function.

C. Weighted U- and V-disparity :

As detailed in section *A*, the likelihood function used for estimating the disparity map is a two steps procedure. We recall that the first step is performed using a rectangular window which gives optimal scores for pixels belonging on obstacles, while the second step uses a sloped window which gives optimal scores for pixels corresponding to the road plane. For each step, a confidence map is obtained by applying the confidence function, detailed in section *B*, in each matched pixel. These confidence maps are termed the *Obstacles Confidence Map (OCM)* and the *Road Confidence Map (RCM)*. Recall that the basic version of the so called V-disparity map is basically obtained by cumulating all of disparity values along lines and columns of the dense disparity map. Each point in the V-disparity space corresponds to the sum of all disparities along the same v image line. Figure 5 illustrates the two cumulating spaces.

According to our knowledge, there is no improvement of the V-disparity method since its inception. In our approach, we propose a more general method which we term *WV-disparity*. It is obtained starting from the dense disparity map, the OCM and the RCM maps. Unlike the standard form of the V-disparity, each pixel in the dense disparity map is weighted by two confidence values as follow $\alpha \times C_o + \beta \times C_r$. The parameters C_o and C_r correspond to the confidence values given by the *OCM* and the *RCM* maps. In our approach, a powerful tool is introduced in the formulation of

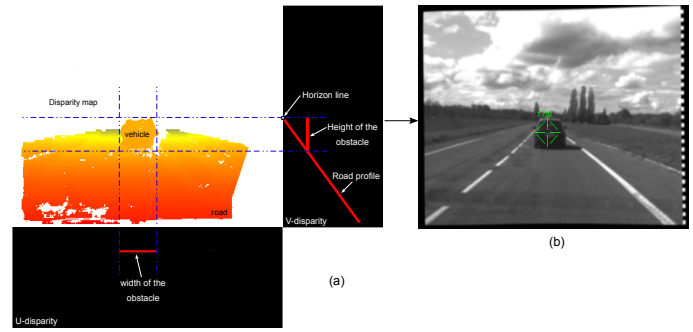


Fig. 5. (a) U- and V-disparity estimation from filtered dense disparity map (b) bounding box of an obstacle in the right image.

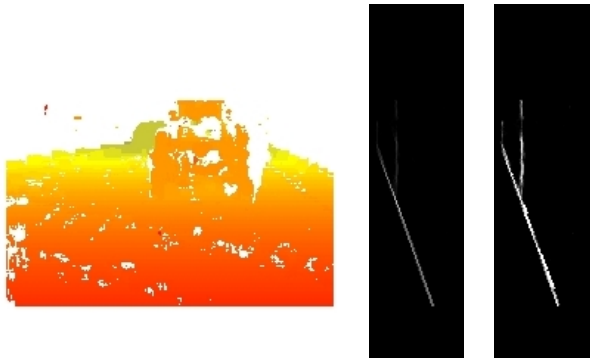
the V-disparity map by re-weighting the confidences values with a non normalized parameters, α and β. Depending on the application, one can give more importance to the obstacles detection component or the road plane subtraction. For obstacles detection, C_o is more weighted than C_r by assigning a hight value to α while a small one for β. This allows highlighting frontal obstacles and ignore the road plane. For road plane extraction, one can simply inverse the weighting of α and β.

Empirically, the obtained weighted cumulative space, i.e. *WC-disparity*, is more accurate than the original V-disparity. The use of the *WV-disparity* and *WU-disparity* maps allows to highlight obstacles lying on the road and best discriminating the most of unlikely obstacles more precisely than using the standard version of the *V-disparity* method. However, the localization of obstacles is done by exploiting the dense disparity map and the cumulative spaces named Weignted U- and V-disparity map. The pixels are clustered into two classes, road and obstacles, after applying a Hough transform on *WV-disparity* and *WU-disparity* in order to highlight each principal line. The presence of vertical line means the presence of a potential obstacle and the presence of a sloped line illustrate the road profile. The v coordinate of the intersection between the road profile and the v-axis gives the horizon line. After the clustering step, the mean weight obtained for all pixels representing each obstacle is computed. This additional information is very useful for the tracking module because of the attribution of an uncertainty measure to each obstacle to be tracked. Figure 6 shows the obtained *WV-disparity* map compared to the standard *V-disparity* map.

In order to identify the most of false alarms, we have defined same filters such as the allowed volume depending on the location of each obstacle. Each filter corresponds to a condition which must be fulfilled and leading to eliminate, if possible, all of erroneous disparities and all of ambiguities regions. A list of obstacles, i.e. target, is therefore obtained on which a tracking procedure is applied.

III. TRACKING ALGORITHM

Once obstacles have been extracted from the two stereo images, a multi-objects association algorithm is needed to estimate the dynamic state of the targets and to monitor



(a) (b) (c)

Fig. 6. The basic *V-disparity* maps vs. the *WV-disparity* map. (a) dense disparity map (b) *V-disparity* map (c) *WV-disparity* map.

appearances and disappearances of tracks. The position of previously perceived objects is predicted at the current time using Kalman Filtering. These predicted objects are already known objects and will be denoted in what follows by Y_j. Perceived objects at the current time will be denoted by X_i. In a general framework, the problem consists in identifying an object designated by a generic variable X among a set of hypotheses Y_i. One of these hypotheses is supposed to be the solution. The current problem consists in associating perceived objects X_i to known objects Y_j. Belief theory allows assessing the veracity of P_i propositions representing the matching of the different objects.

A basic belief allowing the characterization of a proposition must be defined. This basic belief (mass $m_\Theta(.)$) is defined in a [0,1] interval. This mass is very close to the one used in probabilistic approach, except that it is distributed on all the propositions of the referential of definition $2^\Omega = \{A/A \subseteq \Omega\} = \{\varnothing, \{Y_1\}, \{Y_2\}, ..., \{Y_n\}, \{Y_1Y_2\}, ..., \{\Omega\}\}$. This referential is the power set of $\Omega = \{Y_1, Y_2, ..., Y_n\}$ which includes all the admissible hypotheses. These hypotheses must also be exclusive ($Y_i \cap Y_j = \varnothing, \forall i \neq j$). The masses thus defined are called *basic belief assignment* and verify:

$$\sum_{A \subseteq \Omega} m^\Omega(A) = 1 \qquad A \in 2^\Omega, \;\; A \neq \varnothing \tag{6}$$

The sum of these masses is equal to 1 and the mass corresponding to the impossible case $m^\Omega_{1..n}\{X_i\}(\varnothing)$ must be equal to 0. In order to succeed in generalizing the Dempster combination rule and thus reducing its combinatorial complexity, the reference frame of definition is limited with the constraint that a perceived object can be connected with one and only one known object. For example, for a detected object, in order to associate among three known objects, the frame of discernment is $\Omega = \{Y_1, Y_2, Y_3, Y_*\}$ where Y_i means that *X and Y_i are supposed to be the same object*. In order to ensure that the frame of discernment is really exhaustive, a last hypothesis noted Y_* is added. This one can be interpreted as *a target has no association with any of the tracks*. In fact each Y_j represents a local view of the world and the Y_* represents the rest of the world. In this context, Y_* means that *an object is associated with nothing in the local knowledge set*. The tracking module is designed to be a tool allowing a temporal filtering of the set of detected targets. This allows a spatial and temporal filtering of each obstacle. This tracking stage also allows reducing both the false detections rate and the missed detection rate. Previously, this approach has been used in order to track obstacles with telemetric sensors in a cartesian referential [8]. In this work, this approach has been adapted to the image space.

IV. EVALUATION AND RESULTS

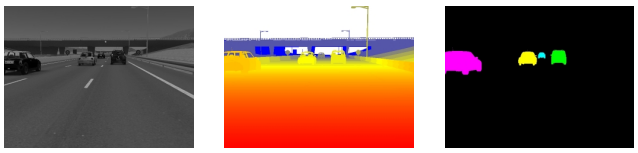
In this section we describe the quantitative and qualitative results of our proposed stereovision-based system for detection and 3D localization of any kind of obstacles in road environment in daytime. We demonstrate the ability of our system to detect obstacles in various geometrical configurations. This evaluation stage has been also done in both simulated and real environments. We begin at the first time by evaluating the performances of our algorithms by referring to a simulated dataset in which all of disparity maps and obstacles ground truth are available [9]. Figure 7 illustrates the Satory's closed loop tracks used for building images database on which the proposed algorithm is evaluated. A lap around the Satory's track is 3.4 km with various geometrical configurations such as straight segments, slight and sharp bends, and a nonconstant road plane, i.e. with different slopes. The track is also equipped with some guardrail. For qualitative evaluation purposes, we need a ground truth at any time in order to get information about the depth of any real point and the set of obstacles. For experiments, the following parameters have been fixed: the baseline, i.e. the distance between the two optical axis of cameras, is 0.5 meter. The size of the optical sensor is of 5.12×3.84 cm which corresponds to an image of size 640×480. For reducing processing time, only the quarter of the image is processed. For the research of corresponding pairs of pixels, the window size is fixed to 8×8. Knowing that the sensors are positioned at 1.4 meter from the road plane, the set of possible disparities values is up to 60. In the following, the ability of the proposed system for detecting obstacles in simulated and real database is demonstrated.

Fig. 7. Satory's track used for evaluationg our algorithm on real environment. A lap around the track is 3.4 km with different geometrical configurations.

A. Evaluation on simulated database

The performance of our algorithm is assessed by comparing the obtained detection results with a ground truth generated by means of the *SiVIC* platform developed in the LIVIC laboratory. This simulator allows modeling any complex environment, sensors, and vehicles with realistic

rendering. However, it offers diverse and useful functionalities such as adding gaussian or arbitrary noise or even rain or fog on simulated images. The database we have generated is obtained by simulating two cameras which gives left and right images with their corresponding ground truth depth maps and obstacles mask. Figure 8 illustrate a right image with its corresponding disparity map and an image on which real obstacles are identified.



(a) (b) (c)

Fig. 8. Images obtained with SIVIC simulator. (a) image acquired from right camera (b) disparity map ground truth (c) list of obstacles.

To better evaluate the localization accuracy, we opted for testing our algorithms with different image sequences and various obstacle configurations. Three simulated sequences had been generated on the virtuel Satory's track. Each one of these dataset contains 5 vehicles with only one moving in front and in the same lane than the ego-vehicle. In the third sequence, the front vehicle dynamically changes its longitudinal position. In the first sequence the frontal obstacle, e.g. a car, is at 10 meters while for the second one is at 23 meters. Note that each sequence has around four thousand frames. For the moment, the accuracy of the 3D localization processes is first performed only on the frontal car. Errors of lateral and longitudinal positions of the frontal obstacle for all database are given in table I.

Frontal obstacle location	**10 m**	**23 m**	**various distances**
Lateral errors	0.11 (m)	0.17 (m)	0.18 (m)
Longitudinal errors	0.34 (m)	0.59 (m)	1.06 (m)

TABLE I

LATERAL AND LONGITUDINAL ERROR POSITIONS (METERS) OF THE FRONTAL OBSTACLE FOR ALL SEQUENCES.

The errors in lateral and longitudinal positions given in Table I are obtained by measuring the Euclidean distance between the centre of mass of both the ground truth and the frontal obstacle. The nearer is the obstacle, the lower is the error. By considering only frontal obstacles, the rate of correct detections for the first two sequences, (i.e. scenarios with constant interdistance: 10 and 23 meters), is 99.52 %. For the last sequence, the rate of correct detections is 95.39 %. The false detections rate is 0 % for all tested sequences. Note that considering all of obstacles in each frame, the correct detection rate are 89.6 %, 94.75 % and 91.2 % for the three evaluated sequences, and the false detection rate is 11.93 %, 12.4 % and 23.31 % respectively. We illustrate in Figures 9 and 10 the errors of lateral and longitudinal positions respectively, given for each frame.

The uncertainty on lateral positions grows significantly either in the case of a turn so that the lateral positions of each obstacle varies substantially, or in the case of partial occlusions due to lane changing of some obstacles.

Another evaluation criterion is expressed by the *Recall* R and *Precision* P parameters which describe the way the obstacles image matches the corresponding ground truth. We remind that the *Recall* measures the ability of an algorithm to well detect obstacles, while the *Precision* is an intrinsic criterion which gives a clue to the accuracy of the detection. Instead of presenting these measures separately, we propose to use a correlated measurement defined by $F_\beta - measure$.

$$F_\beta = \frac{(\beta^2+1)PR}{\beta^2 P + R} \qquad (0 \leq \beta \leq \infty) \tag{7}$$

β is a parameter that controls a balance between P and R. When $\beta = 1$, F_1 comes to be equivalent to the harmonic mean of P and R which equally weighs precision and recall. If $\beta > 1$, F_β becomes more recall-oriented and if $\beta < 1$, it becomes more precision-oriented, e.g., $F_0 = P$.

Figures 11 and 12 shows the ROC curves which express the $F_\beta - measure$ by varying the β parameter for only frontal obstacles and all of obstacles in the observed scene, recpectively. By fixing the β parameter to 1, both of R and P measures are weighted equally. The weight put on R and P critera can be change in order to highlight one of them.

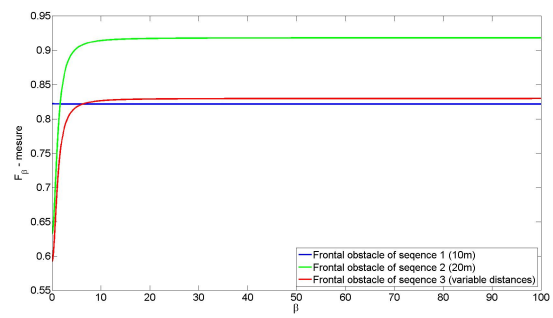


Fig. 11. Evolution of $F_\beta - measure$ by varying the β parameter which attributes less or more weight for *recall* and *precision* measures, for only frontal obstacles of the three sequences.

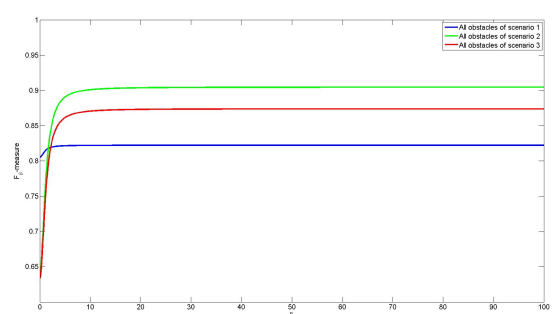


Fig. 12. Evolution of $F_\beta - measure$ by varying the β parameter which attributes less or more weight for *recall* and *precision* measures, for all obstacles of the three sequences.

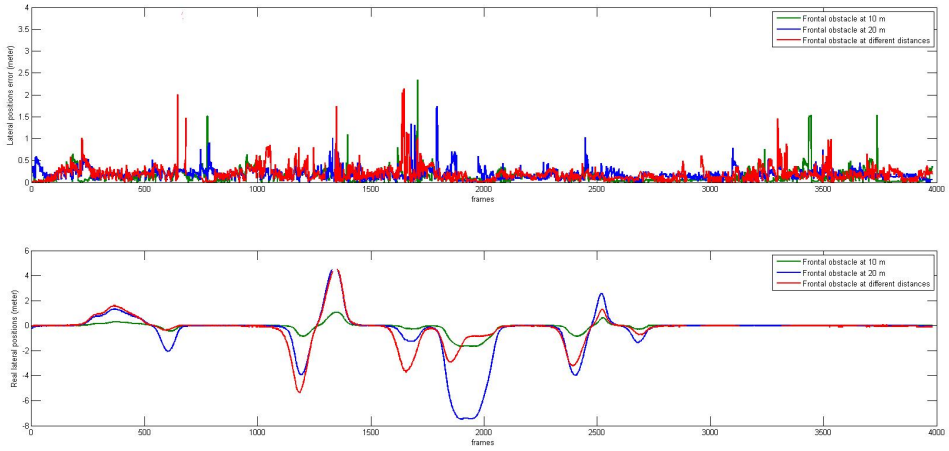


Fig. 9. The error of lateral positions (meter) for frontal obstacles: At 10m (green), 23m (blue) and variable distances (red).

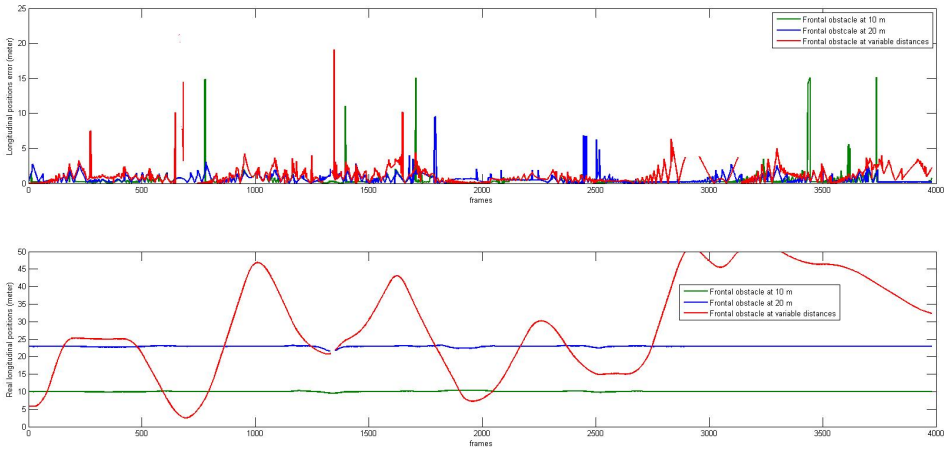


Fig. 10. The error of longitudinal positions (meter) for frontal obstacles: At 10m (green), 20m (blue) and variable distances (red).

(a) (b)

Fig. 13. Overview of the proposed system (a) instrumented vehicle (b) stereovision configuration.

B. Evaluation on real database

An intensive evaluation is made on real sequences acquired in normal and degraded weather conditions such as the case of dazzle by the sun and cloudy time. The vehicle used for experiments is shown in Figure 13 on which we have mounted a pair of stereo cameras with the same geometrical configuration described in the previous section. Some dozen of sequences with four thousand images for each are used for the evaluation.

In the database, the number of obstacles is limited to three vehicles. The evaluated sequences are manually labeled while considering the correct detections, the missed detection, and the false positives rates as a criterion for the evaluation. For each frame, an obstacle is considered as *well detected* if it is highlighted by the stereovision algorithm relatively to the ground truth. The detection range is limited to 5 m for the nearest obstacle and to 55 m for the farthest one. By considering 30000 frames which are processed, the results in terms of correct detections, missed detections, and false detections are given in the following tables:

	Correct detections	Missed detections	Fp
without tracking	96.49%	3.51%	2.74%
with tracking	99.04%	0.96%	0.85%

TABLE II
RATE OF CORRECT DETECTIONS, MISSED DETECTIONS AND FALSE POSITIVES.

Table II shows that the correct detections rate obtained for the real dataset is more important that the one obtained for simulated dataset. This can be explained by the fact that

we do not consider as obstacle both all of obstacles which are partially overlapped by another ones and those which overtake the ego-vehicle on the left and on the right. Figure 14 shows some screenshots of the output of the stereovision detection and tracking algorithm. The Z component of each obstacle, i.e. how far is an obstacle from the ego-vehicle, is displayed over each bounding box shown in green.

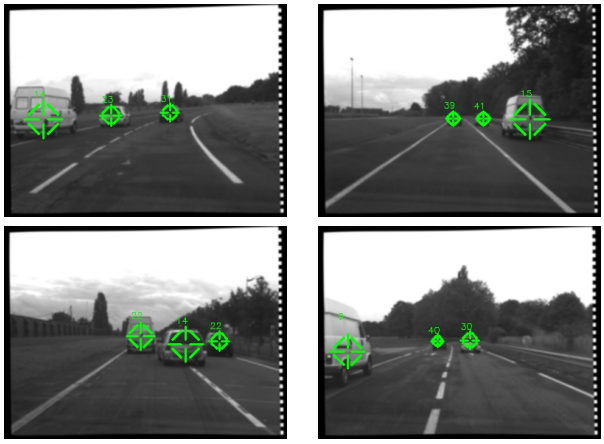


Fig. 14. Screenshot of some detection results obtained from real sequence.

C. Performances analysis

The proposed system has demonstrated the ability to provide a set of relevant obstacles in daytime environment. To be useful and in referring to the main requirement that such a system must be embedded in a vehicle, any system must reach the real-time performances. It is the case with our proposed system which is first evaluated, in term of processing time, on an Intel(R) Xeon 2.00 GHz quadruple core processor. From all the 20 real and simulated videos we have processed, the processing time varied between 30 and 40 ms. This time processing is lower than the sensor sampling time (40 ms).

V. CONCLUSION

We presented comprehensive an efficient obstacle detections, 3D localization and tracking system based on stereovision principle. Results demonstrate the effectiveness of our system in road environment. First, a dense disparity map is estimated starting from two images. A confidence approach is then applied for measuring the matching quality of each pair of matched pixels. Second, a *WU-* and *WV-disparity* maps are built which allows estimating the 2D bounding box for each obstacle. A set of targets is then obtained from which a tracking algorithm is applied. We note that the main contribution of this paper is the extention and the generalization of the *V-disparity* approach by weighting each pixel in the dense disparity map for the estimation of the *WV-disparity* map. Unlike the standard method, the ours offers the possibility of highlighting either the obstacles or the road plane. After the segmentation step, a confidence value is attributed to each obstacle. This additional infomation is very useful for the tracking module which makes it more robust.

The evaluation stage shows a high rate of correct detections whereas a low rate of false detections and missed detections. We have shown that most of false detections are due mainly to the guardrail in front of the ego-vehicle in case of sharp turn which is considered as a potential obstacle. Missed detections mainly occurs in the case that the nearest obstacle from the ego-vehicle is very far away (about 50m). In this situation, the correlation window must be adapted in order to be efficient and to allows a good enough stereo matching. Another cause of failure in detection occurs also in the case that two or more obstacles are at the same distance from the ego-vehicle and are in different traffic lanes. This leads to fuse these obstacles and the algorithm is unable to discriminate objects really present in the disparity map. In order to dealt with this failing, one can make dynamically change the window size for the stereo matching task. The farther the obstacle, the smallest the size of such a window.

In future work and in order to improve the current algorithm, a cooperative fusion stage will be added in order to enhance the accuracy of the obstacles localization. Reducing false negative rate and filtering out false positives can effectively be done by merging the proposed system with a lane markings detection module which allows identifying all of real obstacles lying on the road. Currently, the integration of this algorithm in a dedicated hardware architecture is in progress.

ACKNOWLEDGMENT

The authors would like to thank the eFuture project for funding this work. eFuture is an European IP project of the Seventh Framework Programme (FP7).

REFERENCES

[1] D. Koller, T. Luong and J. Malik, Binocular stereopsis and lane marker flow for vehicle navigation Lateral and longitudinal control, Technical Report UCB/CSD 94-804, University of California at Berkeley, Computer Science Division, 1994.

[2] L. Matthies, *Stereo vision for planetary rovers : Stochastic modeling to near real time implementation*, In International Journal of Computer Vision, Vol. 8, pp. 71–91, 1992.

[3] R. Wang, Y. Xu and Y. Zhao, *A vision-based road edge detection algorithm*, In Proceedings of the IEEE Intelligent Vehicles Symposium, Versailles, France, 2002.

[4] M. Perrolaz, J.D. Yoder, A. Negre, A. Spalanzani and C. Laugier, *A visibility-based approach for occupacy grid computation in disparity space*, In IEEE Trans. on Intelligent Transportation Systems, 2012.

[5] K. Huh, J. Park and D. Hong, *A stereo vision-based obstacle detection system in vehicles*, In Optics and Lasers in Engineering, Vol. 26, No. 2, 2008.

[6] R. Labayrade and D. Aubert, *Robust and fast stereovision based road obstacles detection for driving safety assistance*, In MVA, Nara, Japan, 2002.

[7] N. Fakhfakh, L. Khoudour, J.L. Bruyelle and E. El-Koursi, *Intelligent Survellance System Based on Stereo Vision for Level Crossings Safety Applications*, In Chapter of book 'Recent Developments in Video Surveillance'. ISBN 978-953-51-0468-1, pp. 75-100, 2012.

[8] D. Gruyer and V. Berge-Cherfaoui, *Matching and decision for Vehicle tracking in road situation*, In IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS99, Kyongju, Core, 17-21 octobre 1999.

[9] D. Gruyer, S. Glaser and B. Monnier, *SiVIC, a virtual platform for ADAS and PADAS prototyping, test and evaluation*, In proceeding of FISITA10, Budapest, Hungary, 30 may-4 june 2010.