Evidential occupancy grid mapping with stereo-vision

Chunlei Yu, Véronique Cherfaoui, Philippe Bonnifait

Evidential Occupancy Grid Mapping with Stereo-vision

Chunlei Yu[1,2,3], Veronique Cherfaoui[1,2,3] , Philippe Bonnifait[1,2,3]

***Abstract*—Occupancy grids have shown interesting properties to model the environment for intelligent vehicles perception. In this paper, we present a novel approach to build 2D occupancy grid maps with stereo-vision. Our approach proposes a fitted sensor model based on the disparity space to interpret the stereo-vision information onto an occupancy grid map. The evidential model deals with sensor uncertainties by using Dempster-Shafer theory. Our approach exploits the U-disparity space to model the obstacle information and the V-disparity space to model the road space information. The fusion of these two sources of complementary information results to an enhanced environmental model. In a first data set, experimental results based on real road data and comparisons with Lidar grids show that the proposed evidential sensor model can model efficiently the environment. In a second one, the mapping of a road environment is reported to show the performance of the proposed model with another stereo-vision system.**

I. Introduction

An occupancy grid map [1] is a common tool to model the immediate environment for intelligent vehicles navigation. The map can be built by interpreting sensor readings into evenly distributed cells in the space to represent the presence or absence of obstacles. An occupancy grid is often constructed using range sensor data because a range sensor data provides direct obstacle distance information, which simplifies the modeling and mapping process, as in [2] with an ultrasound, in [3] and [4] with Lidars. In this paper, we present a new approach to build occupancy grid maps using a stereo-vision camera. Building probabilistic occupancy grid maps with stereo camera has been studied by some researchers in the robotic field. [5] has proposed a local occupancy grid mapping using a proposed model in the U-disparity space, and in [6], the authors proposed an approach to build occupancy grids using a dynamic programming method.

When interpreting sensor data into occupancy, uncertainties inevitably arise because of unperceived space and sensors measurements errors. To tackle these problems, Bayesian methods are the foundations of usual frameworks. The information is transformed into probability to take into account uncertainties. The above cited works are based on Bayesian theory except [3], in which the authors proposed to use the theory of Dempster-Shafer [7] to handle uncertainties.

In this paper, we propose a fitted evidential sensor model based on the disparity space to interpret the stereo-vision information onto a 2-dimensional occupancy grid map. We have chosen evidential theory because the proposed sensor model interprets the disparity space of two sources of complementary information: the U-disparity map serves to model the obstacle information, whereas the V-disparity map permits to model the Free space information. Moreover, the framework of the theory of Dempster-Shafer offers powerful tool to merge different sources of complementary information. In this work, we make the hypothesis that the host vehicle runs on a flat road. The road is considered as Free space whereas obstacles are considered as Occupied space.

The paper is organized as follows: section II presents the evidential theory foundations. Section III details the proposed evidential sensor model for the stereo-vision. Section IV shows experimental results based on real road data carried out on two different datasets with different stereo-vision systems. Entropy and specificity are used to compare the information management of the stereo-vision mapping with a classical lidar approach. Finally, conclusions are given is section V.

II. Theory foundations

In the theory of Dempster-Shafer, a frame of discernment Ω is defined to model a specific problem. In the occupancy grid framework, the frame of discernment is defined as: $\Omega = \{O, F\}$, referred as the states of each cell. The power set is defined as $2^{\Omega} = \{\emptyset, F, O, \Omega\}$. For quantitatively supporting the cell states, a mass function (also referred as basic belief assignment BBA) is calculated and provides four beliefs on the state of the cell $[m(F)\, m(O)\, m(\Omega)\, m(\emptyset)]$, where $m(A)$ represents respectively the quantity of evidence that the space is $Free$, $Occupied$, $Unknown$ or $Conflict$. In this work, the proposed sensor model sequentially assigns mass to the cell space for every sensor reading. A powerful application of evidential theory is the fusion of different reliable sources of information. Let m_1 and m_2 be two given mass functions describing the occupancy belief of the same cell. The result of combination using Dempster's fusion rule is: $m_{fusion} = m_1 \oplus m_2$. The normalization process in Dempster's rule has the effect of distributing the belief from the conflict to the other propositions, according to their respective mass. Based on this property, applying Dempster's rule while updating the occupancy grid map with sequential sensor data can provide obstacle free representation of the environment because conflicts is mainly related to mobile objects in the environment or false alarms [4] [8].

III. Evidential sensor model for stereo-vision

Our sensor model takes as input the raw stereo-vision images, and outputs two evidential occupancy grid maps

The authors are with [1]*Sorbonne universités,* [2]*Université de Technologie de Compiègne (UTC),* [3]*CNRS Heudiasyc UMR 7253, France.* E-mail: {chunlei.yu, veronique.cherfaoui, philippe.bonnifait}@hds.utc.fr

modeled as evidence distribution over the space. In [5] the proposed approach needs a pre-processing step to classify the image pixels as road pixels and obstacle pixels, which is still an open problem referred to as road detection. Our approach needs no pre-processing which is an advantage of our model.

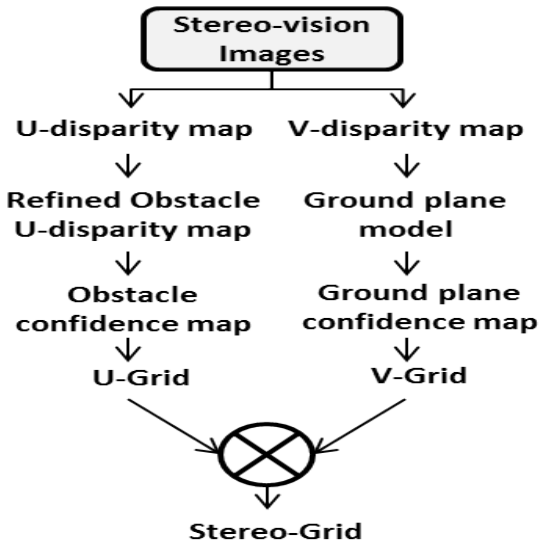


Figure 1: Sensor model overview

A. U-Grid sensor model

When applied to intelligent vehicle applications, U-disparity maps are widely used to detect obstacles [9] [10]. Indeed, with upright obstacle pixel accumulation on the same disparity value one is able to distinguish obstacles from ground. Our approach is similar to this idea. The Occupied mass in the U-Grid comes from the obstacle pixel accumulation. Figure 2 shows the U-disparity computed from stereo-vision images at the bottom. One can notice that the space occupied by upright obstacles contributes to bright cells in the U-disparity map. In our approach, we compute a refined obstacle U-disparity map to compute the $BBAs$ for the U-Grid.

In [11], the authors deduced that the real world height of the objects could be estimated as:

$$h_i = h_c + \frac{(y_i - y_0) * z * \cos\theta}{f}, \tag{1}$$

in which h_i and h_c are the height of object i and camera respectively in the real world coordinate; θ is the camera tilt angle and f is the camera focal length; z is the depth of the object to the camera; y_0 and y_i are the horizon position and top of the objects in image coordinate (the origin of coordinates is assumed at left bottom). In our approach, we push this formula further to calculate the refined obstacle U-disparity map. We transform the equation 1 into:

$$y_i = \frac{f * (h_i - h_c)}{z * \cos\theta} + y_0. \tag{2}$$

Figure 2: Left top: Original left image with calculated horizon line in red; Right top: V-disparity map with calculated ground plane line; Bottom: U-disparity map

Adopting this equation enables to calculate the coordinate of an object in the image coordinates. If one assumes that the whole space in the camera environment is the ground plane, and based on equation 2, one is able to calculate the accumulation value of the ground pixels providing that the disparity value changes from $[d_i - 0.5, d_i + 0.5]$, for each u value in the image coordinate. This process allows to calculate the pure ground U-disparity map for all the ground pixels. If we perform subtraction operation between the U-disparity map and the ground U-disparity map, the resultant U-disparity is defined as refined obstacle U-disparity map, which contains no ground contribution in the U-disparity space.

The Occupied evidence is calculated in the U-disparity space at first. We estimate the accumulation expectation value by defining a detection threshold $[0, H]$ in the real world coordinates. For each cell in the U-disparity map, one can calculate the maximum pixel accumulation expectation V_P by applying Equation 2. Besides in the refined obstacle U-disparity map, the real accumulation value V_O of obstacle pixels at the cell is already calculated. We adopt the ratio between these two values as the obstacle confidence: $\tau_O = V_O/V_P$, and the Obstacle evidence for each cell in the U-disparity space is formulated as: $m_U(O) = 1 - \exp(-\tau_O/\lambda)$, where λ is a constant value. Based on Dempster's theory, the calculated obstacle evidence can only support the degree of belief on the Occupied state and it provides no evidence on the Free state. So the rest of mass besides the $m_U(O)$ is totally assigned to Unknown, so the $m_U(\Omega) = 1 - m_U(O)$.

Please note that obstacle evidence for each cell is in the U-disparity space. We need to transform the information into a Cartesian space. The method is detailed in [5]. We adopt here the same methodology which aims at calculating the influence area in the real world Cartesian coordinates for

each cell $(u,\ d)$ in the U-disparity space.

B. V-Grid sensor model

The Free evidence is modeled as a ground plane confidence map. As discussed in [12], [9], one can deduce the ground plane model based on V-disparity. In Figure 2 the calculated ground plane in the V-disparity space is displayed by the red line. Our approach takes advantage of this property and deduces a ground plane confidence map. This is a pixel-wise map which represents the confidence that each pixel lies on the ground plane. Logically, we suppose that the pixels which possess stronger confidence to be on the ground plane have greater evidence to be Free. Under the hypothesis that the host vehicle moves on a flat road, we can extract the ground plane parameters by applying the Hough transform in the V-disparity space.

In the V-disparity space, the ground plane is modeled as follows:

$$\triangle = a * v + b \tag{3}$$

This is a parametrized line model, as the red line shown in the V-disparity map in Figure 2. $\triangle$ is the expected disparity value so that the pixel can be considered on the ground. To avoid arbitrary decision, we propose to compute the ground confidence map based on the disparity difference between each pixel and the expected disparity value and the confidence is calculated with the Gaussian model. This operation is performed for each pixel line. Algorithm 1 illustrates the ground plane confidence map computation process.

Algorithm 1 Ground plane confidence map calculation

input: disparity map (D), V-disparity map

output: Ground plane confidence map

1. Ground plane extraction based on Hough transform, line parameters a, b.

2. For each image line v_i
   Expected ground pixel disparity: $\triangle_i$ = a*v_i + b
   For each pixel p_{ij} in the line
   Disparity difference $\triangle_d = \triangle_i - D_{ij}$
   Ground confidence for this pixel $C_{ij} = \exp(-\triangle_d^2/\sigma^2)$
   Endfor
   Endfor

The algorithm models the ground confidence of the pixels based on the pixels' disparity distance from expectation. Figure 3 shows the resultant ground plane confidence map projected onto the original left image. The green level in the map illustrates the various confidence degree for each pixel. Only pixels with high confidence are displayed.

Integrating the ground confidence map, the V-Grid takes the confidences as direct support for Free. As a result, the Free mass of a pixel is: $m_V(F) = C_{ij,}\ C_{ij}$ is from Algorithm 1. Since the ground plane confidence map contains no support on obstacle evidence, the rest of mass goes to Unknown, so the $m_V(\Omega) = 1 - C_{ij}$. Note that, as the mass

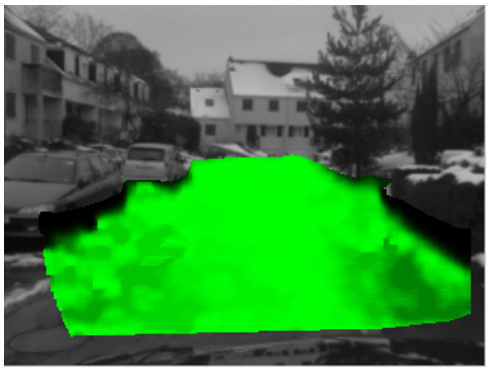


Figure 3: Free evidence projection

distribution is computed for pixels, we need to transform this information into real world Cartesian coordinates. We have adopted the punctual observation model discussed in [13] to assure good accuracy. This model assumes that the imaging process attributes all the probability density to the center of a pixel. From the point of view of a grid cell, the probability is uniform within the cell of the calculated position. This model has the problem of causing discrete evidence information over the space, but a Gaussian model has been adopted for the computation of ground plane confidence, we do not want to add further blurring process into the model. Herein, within the evidential sensor model, information accumulation is based on the Dempster's fusion rule, since there are pixels which correspond to the same cell in the real world coordinates, like the pixels on the bottom of image. By supposing that each of these pixels fall in the same cell as one different source of information, we can apply Dempster's fusion rule.

C. Fusion

Our aim is to build a consistent environmental model based on two complementary sources of information. The two sources of information are both reliable, thus allowing the use of the Dempster's rule. The fusion process is shown in Equation 4. For denotation purpose, let m_U and m_V represent respectively the mass functions of U-Grid and V-Grid at time t. The resultant Stereo-Grid is denoted as m_{Stereo}.

$$m_{Stereo} = m_U \oplus m_V \tag{4}$$

This rule accomplishes the fusion with a normalization process, which can distribute the belief from the conflict to Free and Occupied, according to their respective mass. The complementary U-Grid and V-Grid can be conflicting in some cases. The normalization process of the fusion strengthens the highest belief in the result, which is exactly the desired operation.

IV. Real road experimental results

In this part we report the experimental results based on real road data. The data sets were collected with two experimental vehicles (called Carmen and Zoe) of the Heudiasyc

Laboratory shown in Figure 4a. For comparison purpose, we make use of the Lidar installed in the front of the Carmen. The evidential occupancy grids based on the lidar data were constructed as a source of comparison.

(a) Carmen (b) Zoe

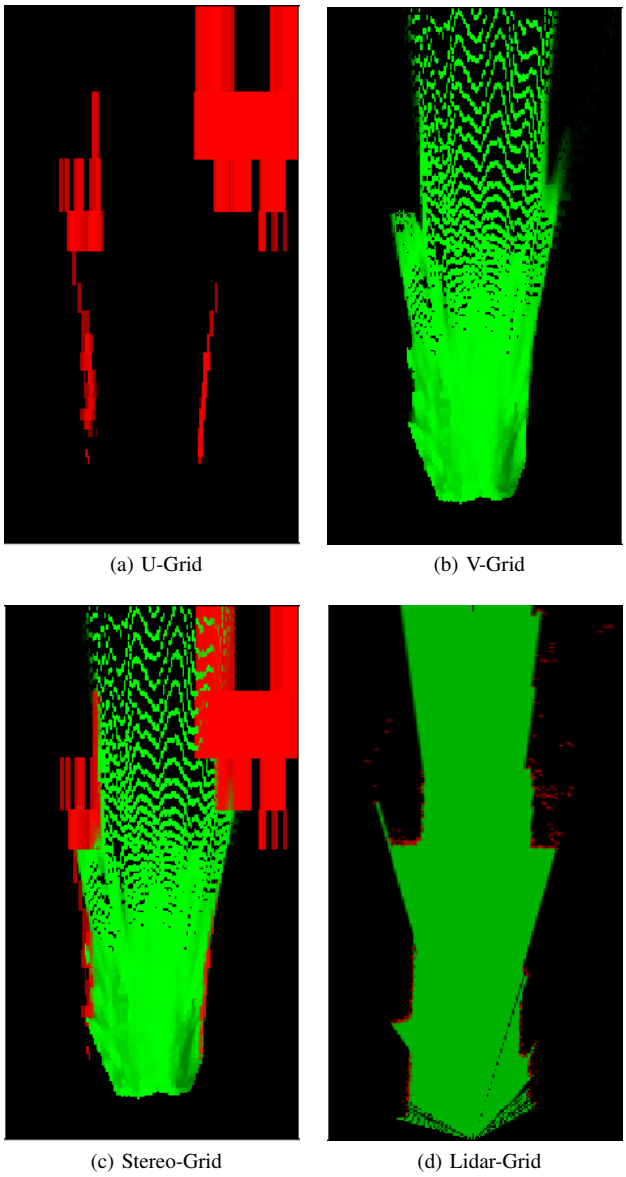
Figure 4: Experimental vehicles used to test the approach

A. Results and comparison with a Lidar-Grid

The results are shown in 15 meters in width $(-7.5,\ 7.5)$, 30 meters in height $(0,\ 30)$ and grid resolution $(0.2,\ 0.2)$ meters. λ was set to 0.1 and σ to 2.5, tuned empirically after several tests. Color brightness level represents the confidence degree. In the U-Grid, brighter color means more obstacle pixels are accumulated, and in the V-Grid, brighter color means that the disparity value of the pixel at this place is closer to the disparity value of the ground pixels. The same rules apply for the Stereo-Grid.

Figure 5 shows the grids constructed from the scene displayed in Figure 2. The car parked on the roadside can be clearly seen in the U-Grid. In the V-Grid, the mass degree variation is clearly noticeable because of different ground confidences and the accumulation effect. In the Stereo-Grid, one can see the fusion results of the two complementary information. The grid reflects both the Free and Occupied information in the environment. Meanwhile, one can remark the space of conflicting information, such as the space under the car and under the tree. These space provides contradictory information in the U-Grid and V-Grid. The normalization operation in the fusion strengthens the one with highest belief. The Lidar-Grid is shown for comparing with the Stereo-Grid. The disadvantage of the Lidar-Grid is the lack of obstacle information due to the sparse nature of the data, but the grid can reflect the environment in a very precise manner. By comparing the Stereo-Grid and the Lidar-Grid, one can conclude that the evidential occupancy grid constructed by the proposed approach is accurate.

Figure 6 shows another scene result. There are parked and on-road cars and a pedestrian on the roadside. From the U-Grid one can see the car on the road and the pedestrian on the roadside are both mapped. The resultant Stereo-Grid correctly shows these information. In the Lidar-Grid, the pedestrian is not obvious to detect due to the small obstacle size. However, in the Stereo-Grid this information is more importantly reflected.



(a) U-Grid (b) V-Grid

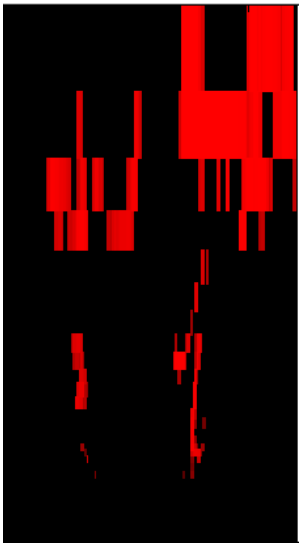
(c) Stereo-Grid (d) Lidar-Grid

Figure 5: Evidential Occupancy Grid for environment with parked car

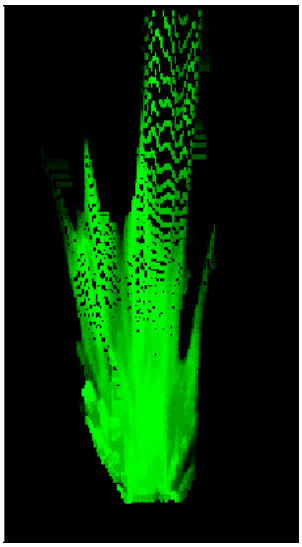
B. Mapping experiments with the Zoe vehicle

In Figure 7, we show a global map constructed by using an different stereo-vision system in a private test site of the university with the robotized Zoe moving at 20 Km/h. The sequential fusion of all the stereo-grids is done with the method presented in [3] which outputs mapping of the area. An image of the test site is also shown in top of the figure. The blue points in the center of the free space represents the host vehicle's trajectory during the data collection. One can notice that the map is correctly built particularly in the roundabout.

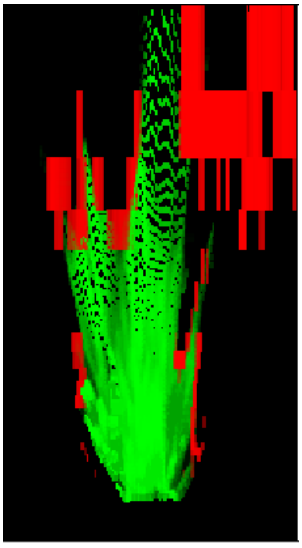
(a) Scene with a running car and a pedestrian



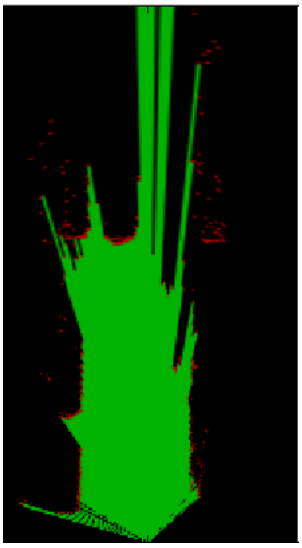
(b) U-Grid



(c) V-Grid



(d) Stereo-Grid



(e) Lidar-Grid

Figure 6: Evidential Occupancy Grid for environment with parked and on-road cars and pedestrain

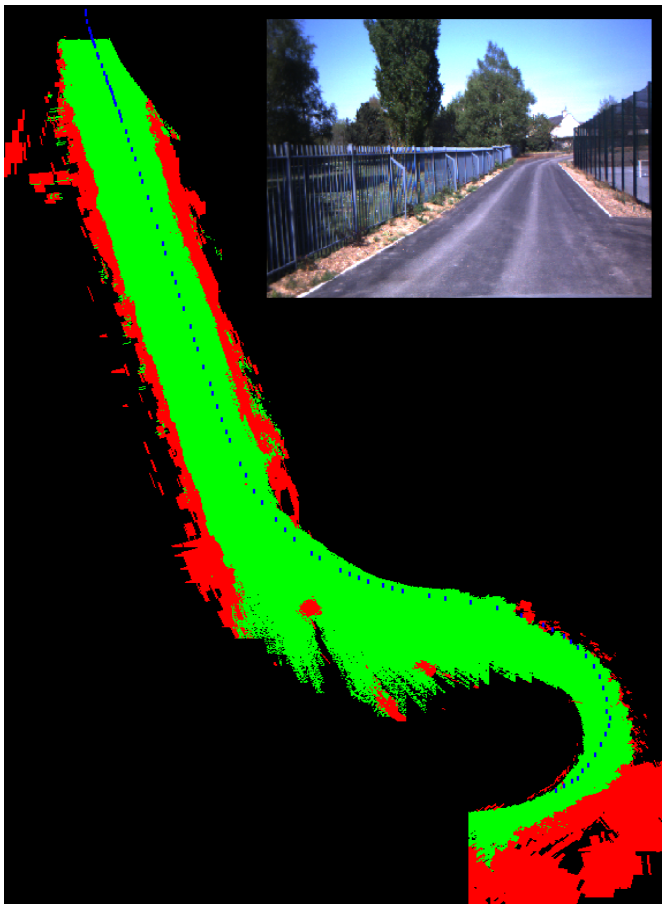


Figure 7: Global map built in test court of Heudiasyc

C. *Quantitative information management study*

In order to better analyze the performance of the stereo-vision evidential mode, we report here a quantitative comparison with the classical lidar approach. In the theory of Dempster-Shafer, one can evaluate the BBAs information using Specificity and Entropy metrics [14]. The entropy of a mass function is defined as follows:

$$E_m = -\sum_{A \subseteq \Omega} m(A) \cdot \ln(\mathrm{pl}(A)), \tag{5}$$

and the specificity of a mass function as:

$$S_m = \sum_{A \subseteq \Omega, A \neq \emptyset} \frac{m(A)}{\mathrm{card}(A)}. \tag{6}$$

An informative and non-ambiguous mass function should have a high degree of specificity and a low degree of entropy. To illustrate these notions of entropy and specificity, please consider the following pedagogical examples with the considered frame of discernment for a given cell:

$$m_1 = \begin{bmatrix} \emptyset & F & O & \Omega \\ 0 & 0.9 & 0 & 0.1 \end{bmatrix} : E_{m_1} = 0 \quad S_{m_1} = 0.95$$

$$m_2 = \begin{bmatrix} \emptyset & F & O & \Omega \\ 0 & 0.1 & 0.1 & 0.8 \end{bmatrix} : E_{m_2} = 0.0211 \quad S_{m_2} = 0.6$$

$$m_3 = \begin{bmatrix} \emptyset & F & O & \Omega \\ 0 & 0.4 & 0.4 & 0.2 \end{bmatrix} : E_{m_3} = 0.4087 \quad S_{m_3} = 0.9$$

Entropy characterizes the inconsistency in the distribution of the masses. No conflicting information results to zero entropy in m_1, while in m_3, as there exists conflicting information between Free and Occupied, the entropy is large. Specificity characterizes the degree of dispersion of the belief. The specificity is larger if the mass distribution is less doubtful, thus in m_1 and m_3 the specificity is large.

We report results for a 20 seconds data sequence carried out with the Carmen vehicle and calculate the average specificity and entropy for every evidential grid. For comparison purpose, stereo-vision and Lidar grids were constructed at the same time index. In Figure 8 and Figure 9, specificity and entropy are displayed. The green and blue points correspond to the Stereo-Grid and Lidar-Grid respectively. In general, the Stereo-Grid has larger specificity than the Lidar-Grid which means that the masses are concentrated on non-vacuous propositions in the Stereo-Grid than in the Lidar-Grid. Thus, less uncertainty about the environment exist in the Stereo-Grid. In other words, the environment is better characterized. This can be explained by the fact that the stereo images contain more information than the sparse Lidar scans. Meanwhile, the Stereo-Grid also has larger entropy than the Lidar-Grid which means the mass is more dissonant in the Stereo-Grid than the Lidar-Grid. This is due to the fact that the Stereo-Grid originates from the fusion of two complementary grids which may contribute to contradictory information in some cells close to the obstacles. It is interesting to notice that the entropy remains very low which indicates a good modeling of the U and V disparity grids. The lidar has almost no entropy because the evidential sensor model has been tailored to avoid this issue.

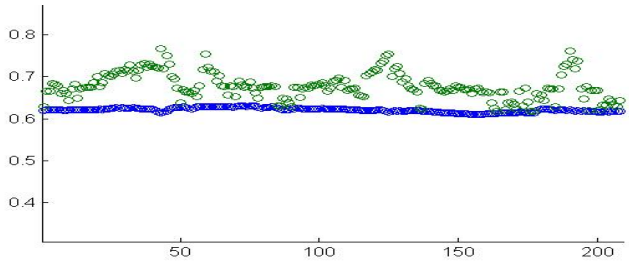


Figure 8: Specificity for different times when the vehicle is moving. x-axis: frame index; y-axis: Average Specificity

V. Conclusion

In this work, an evidential sensor model has been developed for a stereo-vision camera based on disparity maps. The model makes full use of the disparity space and yields two evidential occupancy grids which provide complementary information about the environment. The U-Grid from the U-disparity map contains all information about obstacles in the scene and the V-Grid from V-disparity map models the free space information. The fusion of these two grids offers a complete model of the environment. Experimental results based on real road data with different systems shows that the proposed evidential occupancy grid model can correctly represent the environment, and a quantitative comparison with occupancy grids from Lidar data indicates that the Stereo grids are more informative and less ambiguous than Lidar grids. The main perspectives of this research consist in studying the influence of the parameters used in the stereo-vision model and the information fusion between the stereo-vision and the lidar.

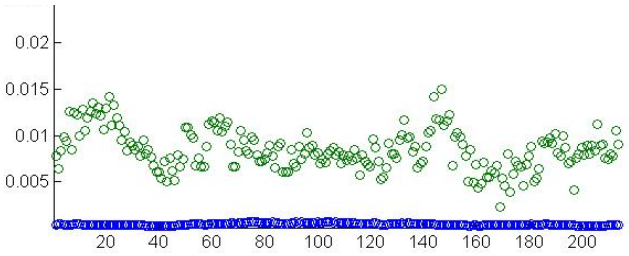


Figure 9: Entropy. x-axis: frame index; y-axis: Average Entropy

References

[1] A. Elfes, "Using occupancy grids for mobile robot perception and navigation," *Computer*, vol. 22, no. 6, pp. 46 – 57, 1989.

[2] S. Thrun, "Learning occupancy grid maps with forward sensor models," *International Conference on Intelligent Robots and Systems, IEEE/RSJ*, p. 6, 2001.

[3] J. Moras, V. Cherfaoui, and P. Bonnifait, "Credibilist occupancy grids for vehicle perception in dynamic environments," *IEEE International Conference on Robotics and Automation*, vol. 6, 2011.

[4] G. Trehard, Z. Alsayed, E. Pollard, B. Bradai, and F. Nashashibi, "Credibilist simultaneous localisation and mapping with a lidar," *IROS-International Conference on Intelligent Robots and Systems*, vol. 8, 2014.

[5] M. Perrollaz, J.-D. Yoder, A. Spalanzani, and Christian, "Using the disparity space to compute occupancy grids from stereo-vision," *The 2010 IEEE/RSJ international conference on intelligent Robotics and Systems*, vol. 6, 2010.

[6] H. Badino, R. Mester, T. Vaudrey, and U. Franke, "Stereo-based free space computation in complex traffic scenarios," *Southwest Symposium on Image Analysis and Interpretation*, vol. 4, 2008.

[7] G. Shafer, *A Mathematical Theory of Evidence*, P. U. Press, Ed. Princeton University Press, 1976.

[8] J. Moras, V. Cherfaoui, and P. Bonnifait, "Moving objects detection by conflict analysis in evidential grids," *IEEE Intelligent Vehicles Symposium (IV)*, vol. 6, 2011.

[9] Z. Hu and K. Uchimura, "U-v-disparity: An efficient algorithm for stereovision based scene analysis," *Intelligent Vehicles Symposium, 2005. Proceedings. IEEE*, vol. 7, 2005.

[10] X. Ai, Y. Gao, J. Rarity, and N. Dahnoun, "Obstacle detection using u-disparity on quadratic road surfaces," *16th International IEEE Annual Conference on Intelligent Transportation Systems*, vol. 6, 2013.

[11] D. Hoiem, A. A. Efros, and M. Hebert, "Putting objects in perspective," *International Journal of Computer Vision*, vol. 13, pp. 3–15, 2008.

[12] R. Labayrade and D. Aubert, "In-vehicle obstacle detection and chracterization by stereo-video," *Int'l Workshop on In-Vehicle Cognetive Computer Vision*, vol. 7, 2003.

[13] M. Perrollaz, A. Spalanzani, and D. Aubert, "Probabilistic representation of the uncertainty of stereo-vision and application to obstacle detection," *IEEE Intelligent Vehicles*, vol. 6, 2010.

[14] R. R. Yager, *Entropy and Specificity in a Mathematical Theory of Evidence*, R. R. Yager and L. Liu, Eds. Springer Berlin Heidelberg, 2008.