HAL
open science

Experiments in Vision-Laser Fusion using the Bayesian Occupancy Filter

John-David Yoder, Mathias Perrollaz, Igor Paromtchik, Yong Mao, Christian Laugier

To cite this version:

John-David Yoder, Mathias Perrollaz, Igor Paromtchik, Yong Mao, Christian Laugier. Experiments in Vision-Laser Fusion using the Bayesian Occupancy Filter. International Symposium on Experimental Robotics, Dec 2010, Delhi, India. inria-00538678

HAL Id: inria-00538678

https://inria.hal.science/inria-00538678v1

Submitted on 23 Dec 2010

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.

Experiments in Vision-Laser Fusion using the Bayesian Occupancy Filter

John-David Yoder[1], Mathias Perrollaz[2], Igor E. Paromtchik[2], Yong Mao[2] and Christian Laugier[2]

[1] Ohio Northern University, Ada, OH, USA. j-yoder @onu.edu
[2] INRIA Grenoble Rhône-Alpes, Saint Ismier, France. Firstname.lastname@inrialpes.fr

Abstract Occupancy Grids have been used to represent the environment for some time. More recently, the Bayesian Occupancy Filter (BOF), which provides both an estimate of likelihood of occupancy of each cell, AND a probabilistic estimate of the velocity of each cell in the grid, has been introduced and patented. This work presents the first experiments in the use of the BOF to fuse data obtained from stereo vision and multiple laser sensors, on an intelligent vehicle platform. The paper describes the experimental platform, the approach to sensor fusion, and shows results from data captured in real traffic situations.

1 Introduction

The primary motivation for this work is moving toward reducing the number of road accidents – which are currently responsible for over a million deaths annually [1]. While autonomous cars such as the entries in the DARPA Urban Challenge [2][3] may increase safety in the very-long-term, in the short and medium-term, technology can hope to reduce the number of accidents through a variety of strategies. As part of the ArosDyn project [4], we are developing an embedded system to provide real-time estimation and prediction of the current level of collision risk. This advance warning of high-risk situations could create enough additional reaction time to let the driver avoid, or reduce the severity of, an accident.

The work presented here is a part of that project, focusing on the experimental challenge of real-time sensor fusion. This work is based on occupancy grids, which have been used to represent the environment [5] and fuse sensor data [6] for quite some time. Stereo vision and laser sensors are both regularly used to create occupancy grids, but unlike other works, this information is fused using the Bayesian Occupancy Filter (BOF) [7]. The BOF, an adaptation of Bayesian filtering to the occupancy grid framework, offers several advantages. Because the BOF is highly parallelizable, it has been shown to run in real-time on a Graphics

Processing Unit (GPU). In addition, the BOF, because it maintains velocity estimates, provides the ability to predict the future motion of dynamic areas in the environment even if they are temporarily obscured. Finally, the output of the BOF is used to provide the estimated position and velocity of objects in the ego-vehicle's environment. Having velocity estimates, rather than just occupancy, allows for more robust clustering [8]. It should be noted that these velocity estimates are relative to the ego-vehicle rather than absolute.

2 Technical Approach

While the overall goal of this work is to improve traffic safety, this paper focuses on an embedded system which can update the driver on the current state of risk in real time. This requires a strong emphasis on efficient algorithms, and on approaches which are highly parallelizable in order to offer good performance on a GPU, and eventually on an embedded parallel processor. The BOF has been specifically developed with the intention of implementing it on hardware for high-performance applications. The BOF also provides the significant advantage of performing sensor fusion at a very low semantic level, the cell level. This reduces the data association problem to its simplest form – there is no need to associate objects or sensor readings directly, only to relate them to a cell in the occupancy grid.

Figure 1 gives an overview of the process. The inputs to the system are a disparity image from the stereo camera system and the laser range data. For each sensor, a probabilistic sensor model converts this input into an occupancy grid. The laser is modeled in a manner similar to that presented in [9]. The stereo-vision sensor uses a novel approach of calculating the occupancy grid first in the u-disparity image [10]. One of the strengths of our u-disparity approach is that it is highly parallelizable, and therefore benefits greatly from implementation on a GPU. It also allows for the detection of partially occluded obstacles, which gives the vision sensor the ability to find objects that the lidar cannot. This is in contrast to other approaches which use vision to construct occupancy grids, such as [11]. In that work, the authors treat the visual sensor very much like a laser, saying that the space before the first object is free, and the space behind the first object is unknown. While this approach allows for simplified processing, it is sensitive to false positives, and does not allow for the detection of partially-occluded objects. In our application, then, this model would remove any advantages of the visual sensor. In essence, there would be no need to fuse sensors, since the visual sensor would not be adding any information compared to the laser.

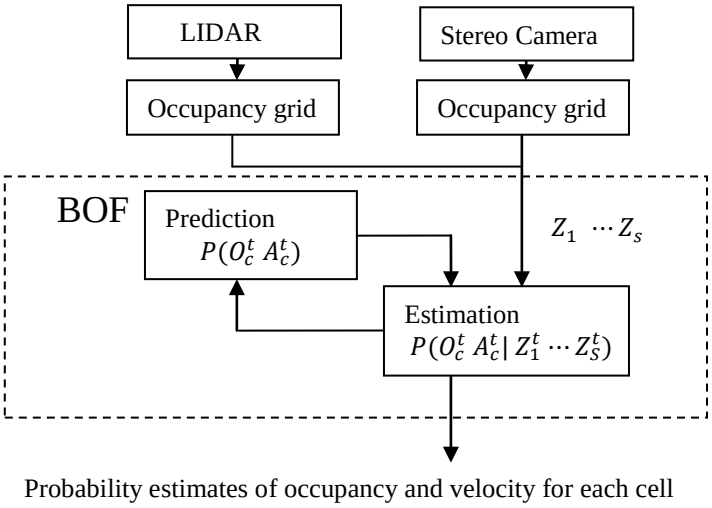


Figure 1: Overview of the sensor fusion process

These occupancy grids (from the laser and from stereo camera) are input into the BOF. As shown in Figure 1, the BOF consists of a prediction phase and an estimate phase. At a given time step t, the prediction phase uses the current probability of occupancy (O) and probability distributions of velocity (A) of each cell to obtain the prediction $P(O_c^t\ A_c^t)$, where c denotes the cell and t the current time step. This prediction is based on the constant-velocity assumption. In the estimation step, $P(O_c^t\ A_c^t)$ is updated based on Z_i^t, the observation of sensor i at time t. This provides the a posteriori state estimate $P(O_c^t\ A_c^t|\ Z_1^t \cdots Z_S^t)$ where S is the number of sensors. Note that in our case the occupancy grid is the same size as the BOF grid, meaning that this occurs for all cells c. This prediction-estimation paradigm provides two major advantages. First, should a sensor reading be unavailable, the BOF continues to produce predictions about the state at each time step. Secondly, it preserves temporal consistency, so suddenly-occluded areas can continue to be tracked for several time steps. Note that the exact number of time steps would depend on the parameters of the BOF and the probability distributions.

The output of the fusion process, then, is the next iteration of the BOF. In our application, the next stage of the process is to model static and dynamic objects in the environment. This is accomplished via the previously published Fast Clustering and Tracking Algorithm [8]. This algorithm provides estimates of the position and velocity of objects in the proximity of the vehicle, along with the associated uncertainties. These values then become input to a risk-assessment model [12].

3 Experimental Platform

The experiments are conducted using a Lexus LS600 as the base for the experimental platform. It was desired to instrument the vehicle for experiments with as little modification to the vehicle as possible. The two laser sensors, IBEO Lux Lidars, are placed in the front bumper. A TYZX stereo camera system is mounted just forward of the rear-view mirror. These modifications can be seen in Figure 2, and are the only modifications visible from the exterior.

Figure 2: Sensors mounted in the Lexus, made available to INRIA by Toyota Europe as part of a long-term ongoing collaboration.

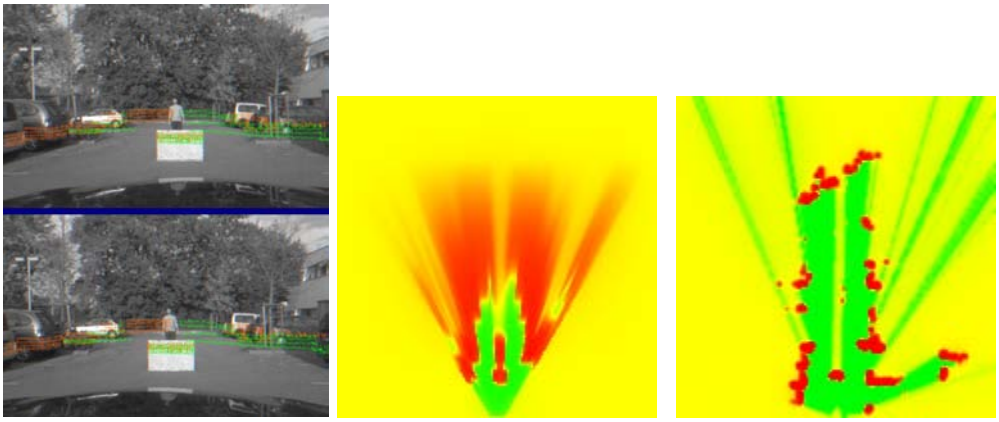
In addition, the vehicle has been equipped with a Dell computer with an NVidia GPU. An Xsens IMU with GPS is placed at the center of the rear axle. Hugr middleware [13] is used to record, synchronize, and replay data sequences. Algorithm parameters (for sensor models, the BOF, etc) are adjustable through a Qt-based user interface.

Specifics for the hardware are as follows. The stereo camera baseline is 22 *cm*, with a field of view of 62°. Camera resolution is 512x320 pixels with a focal length of 410 pixels. Each lidar provides four layers of up to 200 impacts with a sampling period of 20 *ms*. Maximum detection range is about 200 *m*, the angular range is 100°, and the angular resolution is 0.5°. Using the two lidars, the observed region is 40 *m* long by 40 *m* wide, with a maximum height of 2 *m*. Cell size for the occupancy grids is 0.2x0.2 *m*. It should be noted that each layer of laser measurements is treated as a different sensor. This means that the BOF is really fusing 8 laser measurements with the stereo vision measurement. The on-board computer is equipped with 8GB of RAM and an Intel Xeon 3.4GHz processor. The GPU is an NVIDIA GeForce GTX 480.

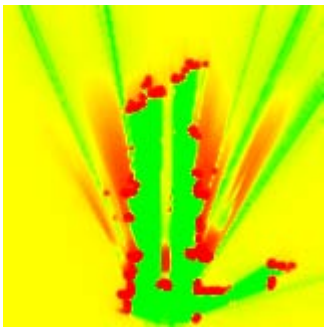
The datasets described in this paper were obtained on French roadways. Further processing, including sensor fusion, was done off-line. Determination of the extrinsic parameters of the sensors was done manually for this work.

4 Results

Figure 3 shows the advantage of sensor fusion, the ability to find a partially occluded object (a pedestrian in a parking lot). Figure 3a shows the view from both cameras, with impacts from the lidar represented by colored dots in the images. Red and green dots denote hits from the left and right lidars, respectively. Figure 3b shows the occupancy grid created by a BOF based on only the stereo camera. Figure 3c shows the occupancy grid created by a BOF based on only the lidars. Finally, Figure 3d shows the occupancy grid obtained from the fused BOF.



a b c



d

Figure 3: Example of Sensor Fusion: (a) image of the scene, (b) occupancy grid created by the stereo vision, c) occupancy grid from the lidar, and (d) fused grid with the BOF.

In Figure 3b, we see the occupancy grid created by the stereo vision sensor. Green cells correspond to low likelihood of occupancy (close to 0), while bright red cells correspond to high probabilities of occupancy (close to 1). The background yellow color represents the case of unknown occupancy (value of 0.5). The stereo sensor has detected the road in the foreground (unoccupied cells), ob-

stacles along the perimeter (parked cars), and the box and person in the center of the image. This frame shows the ability of the visual sensor to detect the partially-occluded pedestrian. Figure 3c shows the occupancy grid from lidar. As is typical of laser sensor models, the area in front of the obstacles has a low likelihood of occupancy, whereas areas behind the obstacles are unknown. As such, the box in the center of the image is found, but the pedestrian is not. The fused image in Figure 3d maintains the advantage of both sensors, finding the pedestrian, and maintaining the more defined obstacles around the perimeter.

The two calculation-intensive portions of the algorithm are the visual processing and the updating of the BOF. As described earlier, one of the central objectives in the creation of the BOF was that its highly parallel structure could allow real-time processing on a GPU, and eventually on an embedded processor. Similarly, the u-disparity approach for occupancy grid creation from the stereo cameras is highly parallelizable. To give some sense for this, the following processor times are provided. We are using a grid size of $40m$ x $40m$ with a cell size of $0.2m$ x$0.2m$. For the computer vision algorithms on the CPU, the computation takes 159 ms, or about 6 Hz, too slow for real-time operation. However, when implemented in CUDA on the GPU, this same algorithm runs in 8 ms. On the GPU, the BOF updates require 5 ms, including the sensor fusion. Thus the stereo processing and the BOF updates can readily be accomplished in real time.

Figures 4-6 show example situations based on data sets obtained urban traffic. In each case, there are five images: (a) shows the scene as viewed by the left camera, (b) shows the occupancy estimates from the BOF, (c) and (d) are the occupancy grids created by the lidar for input to the BOF, and (e) is the occupancy grid created by stereo vision for input to the BOF. Note that the color scheme for the BOF output (b) is as described above. For the input occupancy grids, black represents open areas (close to 0), white represents occupied areas (close to 1), and the background color of gray represents the case of unknown occupancy (value of 0.5).

First, consider the case of Figure 4. The scene in (a) is quite cluttered with a variety of static and dynamic obstacles (pedestrians, cars, a bus and buildings in the background, etc.) The two pedestrians in the foreground can be seen as two small red circles at the bottom of the BOF output in (b). The car, the buildings at the corners of the intersections, etc., all appear in the BOF, as does the pedestrian crossing on the other side of the intersection. When inspecting the occupancy grids (c-e), we can see that the stereo vision produces a less certain occupancy grid (there are more shades of gray). This is due to the novel method of creating this grid. As such, in areas where the lidars are confident, the vision adds little information. In the background, where the lidar has low certainty, the information from vision can “fill in” the grid.

Figure 5 presents us with a good example of this. This scene is shortly after the scene from Figure 4, the main difference being that the pedestrians have moved out of the scene and the bus has moved closer. Inspection of the BOF in Figure 5b shows that, with a fair degree of certainty, the bus is being observed (the long red

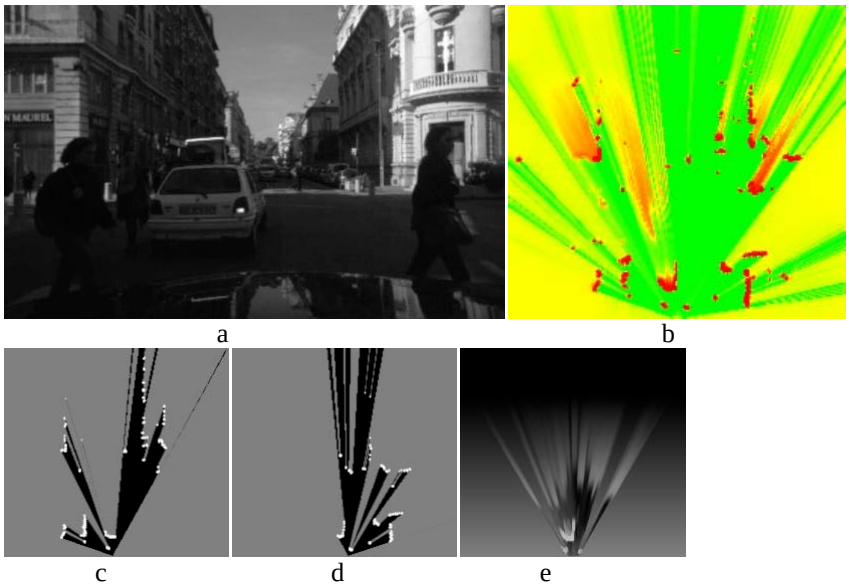


Figure 4: Example Scene. (a) view from left camera. (b) occupancy based on BOF. (c,d) occupancy from left and right lidar. (e) occupancy from stereo

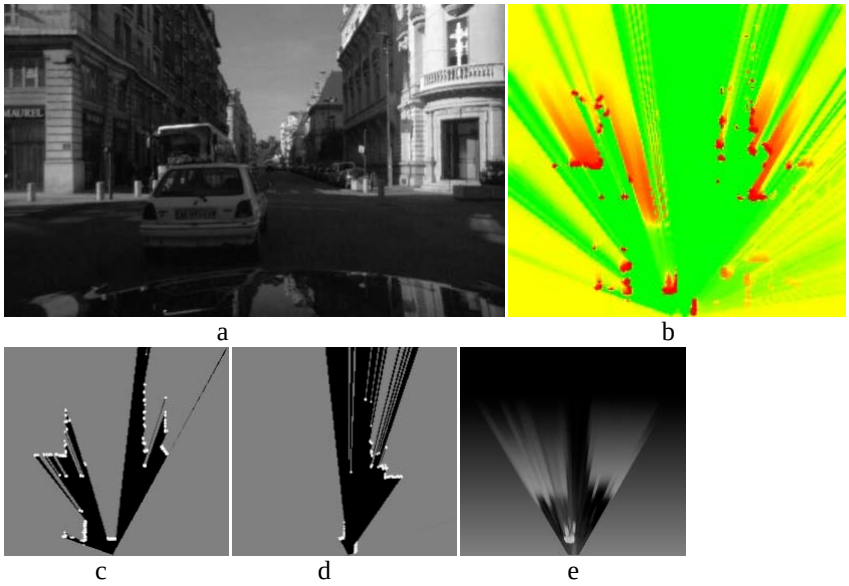


Figure 5: Example Scene. (a) view from left camera. (b) occupancy based on BOF. (c,d) occupancy from left and right lidar. (e) occupancy from stereo

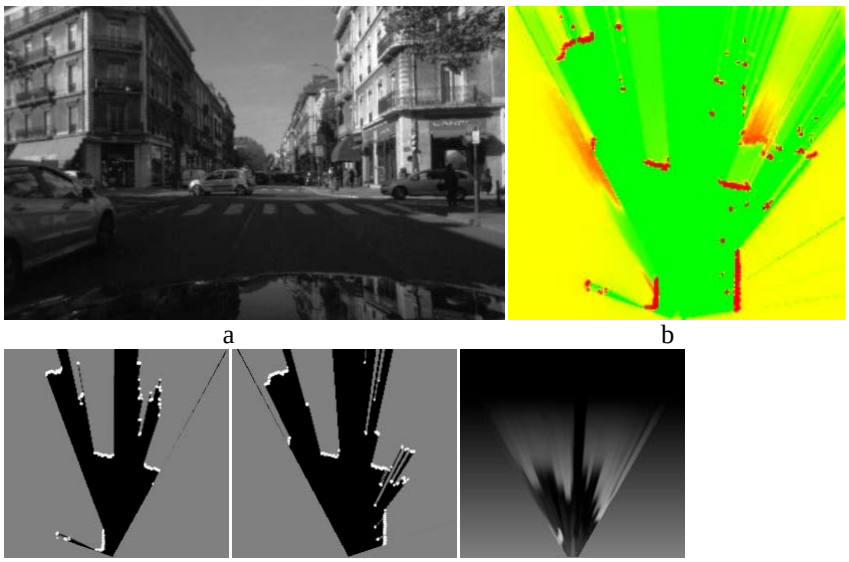


Figure 6: Example Scene. (a) view from left camera. (b) occupancy based on BOF. (c,d) occupancy from left and right lidar. (e) occupancy from stereo

area just to the left of center). However, looking at the grids from lidar (c,d), the bus is not observed – the car in the intersection is blocking the "view" of the lidar. The stereo vision, however, can see over the car to detect the bus. As such, that occupancy grid shows both the car (towards the bottom) and the bus (a bit further up, though it blends in with the building). So in this case, the vision has allowed the BOF to "find" a partially occluded object, in much the way it found the person if Figure 3.

Finally, Figure 6 shows another advantage of the BOF. In this case, the two cars in (a) crossing the intersection are occluding part of the scene, and a pedestrian is occluding part of the car to the right. However, because of the prediction-estimation framework of the BOF, cells are still estimated to be occupied behind the cars (notably the building on the right) despite the fact that they are not in any of the occupancy grids at this instant.

5 Conclusion

The paper describes initial experiments with vision-laser fusion using the BOF in an intelligent vehicle application. The BOF takes as input occupancy grids based on probabilistic sensor models of the laser and stereo-camera sensors. The stereo-camera model has been developed for this application, specifically to allow the detection of partially occluded obstacles. The BOF and the stereo-camera model have been developed to be highly parallelizable, and have been shown to run on a GPU at speeds allowing real-time operation. It is also important to note that since the BOF includes velocity estimates as well as occupancy information, objects can be tracked even while they are temporarily hidden.

The paper has described the sensor-equipped Lexus which serves as the experimental platform for this work. This vehicle has recently been driven in city traffic situations to begin gathering road data. Our experiments have shown that the BOF results in efficient and robust fusion of sensor data from stereo vision and lidars. On a GPU, this can be run in real time. Work is ongoing to test this approach with additional urban driving data, and to evaluate the sensitivity to various BOF parameters. The output of the BOF is being used to identify static and dynamic objects, which in turn will be used to provide real-time assessment of the collision risk as part of the ArosDyn project at INRIA. Additionally, work is ongoing to automatically calibrate the extrinsic parameters of the stereo camera and laser sensors.

Acknowledgements

The authors wish to thank Toyota Motors Europe for ongoing support of this experimental work on the Lexus car. ProBayes is a partner with INRIA Grenoble Rhône-Alpes in the development and patenting of the BOF. Further thanks to Amaury Nègre, Nicolas Turro and Jean-Francois Cuniberto (INRIA) for their technical assistance in setting up our experimental platform.

References

[1] World Health Organization, "Global Status Report on Road Safety: Time for Action". http://whqlibdoc.who.int/publications/2009/9789241563840_eng.pdf

[2] C. Urmson et al. "Autonomous driving in urban environments: Boss and the urban challenge". *J. of Field Robotics*, vol. 25(8), 2008.

[3] M. Montemerlo et al. "Junior: The Stanford entry in the Urban Challenge". *J. of Field Robotics*, vol. 25 (9), 2008.

[4] I.E. Parmotchik, C. Laugier, M. Perrollaz, M. Yong, A. Nègre, C. Tay, "The ArosDyn Project: Robust Analysis of Dynamic Scenes". Proc. of the *11th Int. Conf. on Control, Automation, Robotics, and Vision*, Singapore, 2010.

[5] H.P. Moravec, A.E. Elfes. "High Resolution Maps from Wide Angle Sonar". Proc. of the *1985 IEEE Int. Conference on Robotics and Automation*, pp.116–121, March 1985.

[6] H.P. Moravec, "Sensor Fusion in Certainty Grids for Mobile Robots". *AI Magazine*, 9(2), 1988.

[7] C. Coue , C. Pradalier, C. Laugier, T. Fraichard, P. Bessiere. "Bayesian Occupancy Filtering for Multitarget Tracking: An Automotive Application". *Int. J. Robotics Research*, No. 1, 2006.

[8] K. Mekhnacha, Y. Mao, D. Raulo, C. Laugier. "Bayesian Occupancy Filter based "Fast Clustering-Tracking" Algorithm," *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, Nice, 2008.

[9] S. Thrun, W. Burgard, D. Fox, "Probabilistic Robotics", MIT Press, 2005.

[10] M. Perrollaz, J.-D. Yoder, C. Laugier, "Using Obstacle and Road Pixels in the Disparity Space Computation of Stereo-vision based Occupancy Grids," *Proc. of the IEEE Int. Conf. on Intelligent Transportation Systems*, Madeira, Portugal, 2010.

[11] D. Murray, J. Little, "Using real-time stereo vision for mobile robot navigation," *Autonomous Robots*, vol. 8, January 2000.

[12] C. Tay. "Analysis of Dynamics Scenes: Application to Driving Assistance". PhD Thesis, INRIA, France, 2009.

[13] CyCab Toolkit, http://cycabtk.gforge.inria.fr/