Low level data fusion of laser and monocular color camera using Occupancy Grid framework

Qadeer Baig
University of Grenoble1 & INRIA Rhône-Alpes
Grenoble, France
Qadeer.Baig@inrialpes.fr

Olivier Aycard
University of Grenoble1 & INRIA Rhône-Alpes
Grenoble, France
Olivier.Aycard@inrialpes.fr

***Abstract*—In this paper we have developed a technique for low level data fusion between laser and monocular color camera using occupancy grid framework in the context of internal representation of external environment for object detection. Based on a small variant of background subtraction technique we construct an occupancy grid for camera and fuse it with the one constructed for laser to get a combined view. The results obtained using Cycab simulator prepared by INRIA show the effectiveness of our technique.**

***Index Terms*—Laser-Vision Data Fusion, Occupancy Grids, Robot Environment Representation.**

I. Introduction

Perceiving or understanding the environment is a very important step in the design of autonomous vehicles. To solve this problem, autonomous vehicles are equipped with different sensors in order to perceive their environment and monitor the execution of the planned motion. To be useable, raw data provided need to be processed. An important step in this processing is the fusion of data coming from different sensors. Fusion is the process of combining information from multiple sources in such a way that the combined information are more useful in some sense.

In many applications, to perform fusion, a geometric point of view is used: a set of geometric features is first defined, a model of uncertainty associated to each feature is also needed and a way to fuse features has also to be provided. For instance, [1] used infrared camera and radar to detect and track road obstacles. Each sensor returns a point as observations of the position of each obstacle present in the environment. The uncertainty associated to this position is modelled by a gaussian and when two observations correspond to the same obstacle a fusion of the two corresponding gaussians is performed to estimate the position of the object. In [2], a generalized feature model for the multi sensor case has been developed. This generalized feature model is based on the assumption that any entity in the world can be detected and recognized by means of features. Features are assumed to be dedicated parts of the entity with certain spatio-temporal coordinates in the coordinate system of the entity. Actually, the major drawback of the geometric approach is the number of different geometric features (points, segments, polygons, ellipses, etc) that the perception system must handle. Moreover, this approach is unable to take into account a new objet that appears in the environment and that could not be defined using the predefined set of features.

An other way to model the environment has been introduced by Elfes and Moravec at the end of the 1980s. This framework to multi-sensor fusion is called Occupancy Grids (OG). An occupancy grid is a stochastic tessellated representation of spatial information that maintains probabilistic estimates of the occupancy state of each cell in a lattice [3]. The main advantage of this approach is the ability to integrate several sensors in the same framework taking the inherent uncertainty of each sensor reading into account, in opposite to the Geometric Paradigm whose method is to categorize the world features into a set of geometric primitives. The alternative that OGs offer is a regular sampling of the space occupancy, that is a very generic system of space representation when no knowledge about the shapes of the environment is available. The occupancy grid paradigm has been applied successfully in many different ways. For example, some systems use occupancy grids to plan collision-free paths [4] or for path planning and navigation [5] [6]. Therefore, most of actual mapping systems resort to OG for modelling the environment [5], [7]. And above all with appropriate sensor models OG provides a rigorous way to manage occlusions in the sensor field of view. Contrary to the feature based environment model, the only requirement for an OG building is a bayesian sensor model. This sensor model is the description of the probabilistic relation that links sensor measurement to space state, that OG necessitates to make the sensor integration.

In this paper, we present a new method to perform fusion between two kinds of sensors: a laser scanner and a monocular color camera. In next section, we present the experimental platform used to evaluate the solution we propose. Section III defines the occupancy grid's basic concepts. In section IV and V, we detail how the sensor model of laser scanner and of camera are built. Fusion process is detailed in section VI. Experimental results are reported in section VII. We give some conclusions and perspectives in section VIII.

II. Experimental Setup

The demonstrator vehicle used to get data sets for this work is a simulator of a Cycab[1] vehicle prepared by INRIA. This simulator provides a graphical interface along with a movable

[1] http://cycabtk.gforge.inria.fr/wiki/doku.php?id=download

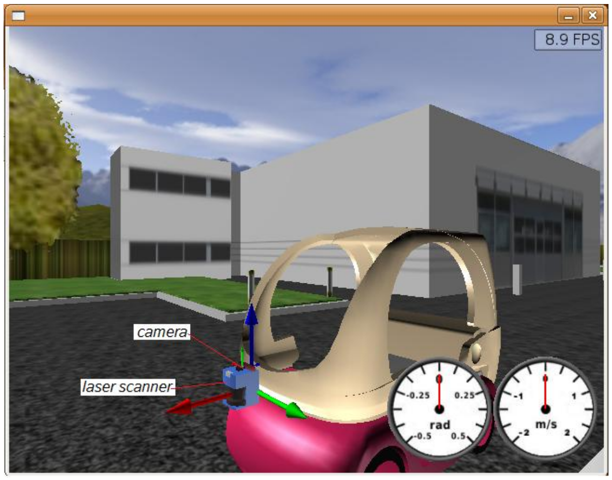

Fig. 1. Cycab with laser scanner and camera mounted on it.

Cycab car. This Cycab car has simulated laser and camera sensors fixed on it. We can also load different static and moving objects in the view by executing their respective scripts. With this simulator creation of experimental environment becomes very easy.

The figure 1 provides a close view of the Cycab vehicle and the two sensors mounted on it. In the simulator environment this Cycab vehicle can move around in the parking.

The simulated laser scanner has a field of view of $180°$ and a maximum range of 90 meters. It has 360 laser beams called channels.

III. Occupancy Grid

In this section we will describe the process of constructing occupancy grid for generic sensor's data, but first a word on notations. We denote the discrete time index by the variable t, the sensor observation from vehicle at time t by the variable z_t where z is a vector of $n >= 1$ dimensions that depends on the sensor used.

Occupancy grid is a multi-dimensional tessellation of space into cells where each cell stores a probabilistic estimate of its state. In this representation, the vehicle environment is divided into a two-dimensional lattice M of rectangular cells and to each cell M_i is associated an occupancy probability having a value between $[0, 1]$ indicating whether a cell is occupied or free. A high value of cell probability indicates the cell is occupied and a low value means the cell is free. Here we make the assumption that occupancy states of individual grid cells are independent. At an instant of time t the state of a cell M_i can be estimated by the posterior probability of occupancy $P(M \mid z_{1:t})$ but because of cell independence assumption it is equivalent to estimate $P(M_i \mid z_{1:t})$ for each cell M_i of grid M, given observations $z_{1:t} = \{z_1, ..., z_t\}$.

In the literature, many methods are used for occupancy grid mapping, such as Bayesian [8], Dempster-Shafer [9] and Fuzzy Logic [10]. Here we apply Bayesian Update scheme [11] that provides an elegant recursive formula to update the posterior for new observations under log-odds form:

$$\begin{aligned} log\,Odds(M_i \mid z_{1:t}) &= log\,Odds(M_i \mid z_{1:t-1}) \\ &\quad + log\,Odds(M_i \mid z_t) - log\,Odds(M_i) \qquad (1) \\ \text{where } Odds(a \mid b) &= P(a \mid b) \,/\, (1 - P(a \mid b)) \end{aligned}$$

In equation (1) $P(M_i)$ is the prior occupancy probability of the map which is set to 0.5 representing an unknown state. The remaining probability $P(M_i \mid z_t)$, is called the *inverse sensor model*. It specifies the probability that a grid cell M_i is occupied based on a single sensor measurement z_t. Moreover, since the updating algorithm is recursive, it allows for incremental cell state updating when new sensor data arrives. To construct an occupancy grid for a sensor we must define an appropriate inverse sensor model for that sensor.

It is easy to see that the desired probability of occupancy, $P(M_i \mid z_{1:t})$, can be recovered from the log-odds representation using the equation (2).

$$P(M_i \mid z_{1:t}) = 1 - \frac{1}{1 + \exp log\,Odds(M_i \mid z_{1:t})} \qquad (2)$$

IV. Occupancy Grid for Laser

For laser sensor the measurement is defined as $z_t = \{z_t^1, ..., z_t^K\}$ for K individual measurements corresponding to K laser beams. A laser beam reading looks like $z = [r, \theta]^T$ where r is the distance and θ is the beam angle. In order to construct occupancy grid for laser we need to define the inverse sensor model corresponding to $P(M_i \mid z_t)$ term in equation 1. Figure 2 shows a well known inverse sensor model for laser sensor. This corresponds to detecting an object at distance d, so a high value of probability near d, the width of this almost bell shaped curve near d pertains to the uncertainty of the laser sensor. The flat curve of low probability before the high value means space is empty between the sensor and the object. A constant value of 0.5 probability after the distance d means that the occupancy state of the cells is unknown past the object at distance d because this object is hiding anything behind it.

In our implementation of this inverse sensor model, the occupancy state of a cell is decided by the nearest beam from the center of this cell. If the nearest beam goes beyond this cell it has low probability of occupancy. But if the nearest beam terminates in the cell or quite near this cell, it has high value of occupancy. Otherwise cell has a value of 0.5 indicating that the occupancy state is unknown.

V. Vision Processing

As we know that an image from a camera is a 2D projection of a 3D view. Whereas a laser scanner scans semi-circle shaped horizontal plane at a fixed height from ground. Since our target is to construct two grids, one for laser and one for camera, and then to merge them to get a combined occupancy state of the view. Therefore it is very important that constructed grids correspond to the same area of the 3D view for both sensors. Horizontal plan scanned by laser corresponds to a sort of horizontal strip in the image. This implies that if we want to calculate occupancy grid for camera then we, first of all, need to find the points in the image corresponding to each

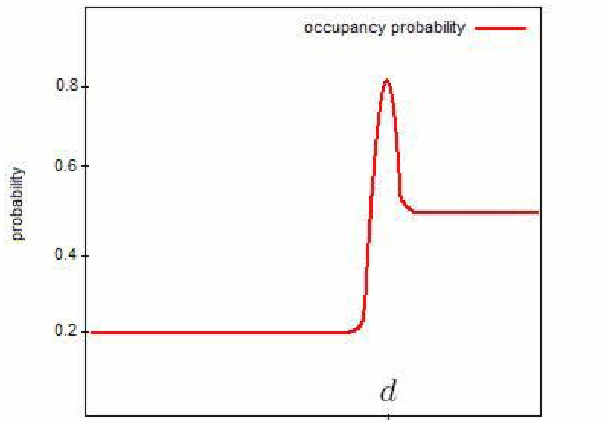

Fig. 2. Probability curve of inverse sensor model for a laser sensor detecting an object at distance d.

laser beam. This requires us to perform transformations from laser reference frame towards camera reference frame. These transformations will give us pixels in the image that correspond to the hit points of the laser beams. So for each laser hit point $P = [x, y, z, 1]^T$ we need the pixel $X = [u, v, 1]^T$. For laser we have $x = r\cos(\Theta)$,$y = r\sin(\Theta)$ and $z = 0$ since our global origin is same as that of laser reference frame's origin. We use following equations from vision literature to calculate these transformations.

$$X_{3\times1} \sim PM_{3\times4} \times P_{4\times1} \tag{3}$$

Where Projection Matrix PM is given as:

$$PM_{3\times4} = K_{3\times4} \times \begin{pmatrix} R & -Rt \\ 0^T & 1 \end{pmatrix}_{4\times4} \tag{4}$$

Here K is the camera calibration matrix. R is rotation matrix and t is the camera translation vector. Camera is calibrated and fixed with respect to laser scanner so we know values of these matrices and vector.

By applying these equations we can calculate image pixels for each laser beam if its termination point lies in the camera view. Now in order to classify these pixels as belonging to objects or background, we need a technique to differentiate background from the foreground. This technique is explained below.

A. Object Detection

Our object detection technique is based on background subtraction. We classify a pixel as either belonging to background or foreground. In order to apply this technique we need to learn background first. We have used multivariate Gaussian distribution technique to learn background. In our case bounding fence of the parking area forms the background. Learning background essentially means learning mean vector (μ) and covariance matrix(Σ) of the multivariate Gaussian distribution. For this purpose we collect different sets of pixels belonging to background from different distances by moving cycab in the parking area and manually segmenting the regions belonging

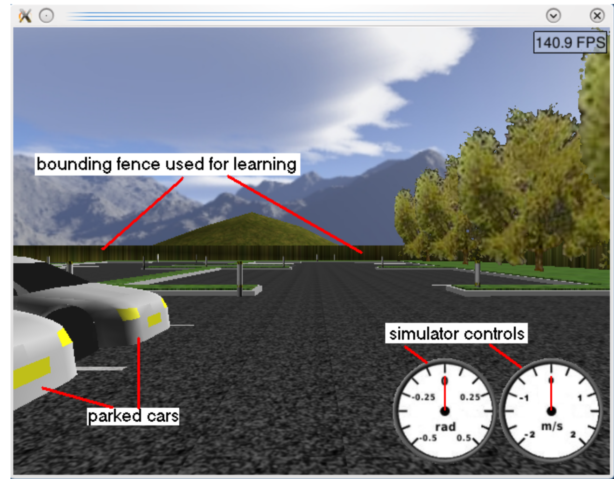

Fig. 3. Simulated parking area.

to background. Figure 3 shows the simulated parking area with bounding fence and parked cars in it. Some example background segments belonging to the bounding fence are shown in the figure 4. Now we can calculate the μ and Σ parameters of the multivariate Gaussian distribution for N background pixels $X = [r, g, b]^T$ as follows:

$$\mu = \frac{1}{n}\sum_{i=1}^{N} X_i \tag{5}$$

$$\Sigma = \frac{1}{n}\sum_{i=1}^{N} (X_i - \mu).(X_i - \mu)^\top \tag{6}$$

The advantage of this technique is that object detection will work even if the simulator cycab is moving in the parking area.

After we have learnt background parameters we can apply following equation to classify any pixel $x = [r, g, b]^T$ as belonging to object or background.

$$f_x(r, g, b) = \frac{1}{(2\pi)^{3/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right) \tag{7}$$

If the probability is smaller than a threshold value, learnt experimentally, then the pixel belongs to an object otherwise it belongs to background.

B. Occupancy Grid For Camera

For camera measurement is defined as $z_t = \{z_t^1, ..., z_t^k\}$ for k individual pixels corresponding to k laser beams' termination points visible in the camera view. Whereas one pixel measurement is defined as $z = [c]$ where c is the class of the pixel. For constructing an occupancy grid we need to define inverse sensor model for camera to calculate $P(M_i \,|\, z_t)$ term in equation 1. We define inverse sensor model as follows, if a pixel is classified as belonging to background then all the cells lying along a line from camera to the end of the grid have low value of occupancy. But if the pixel is classified as objet then all the cells lying along the line have relatively high value of

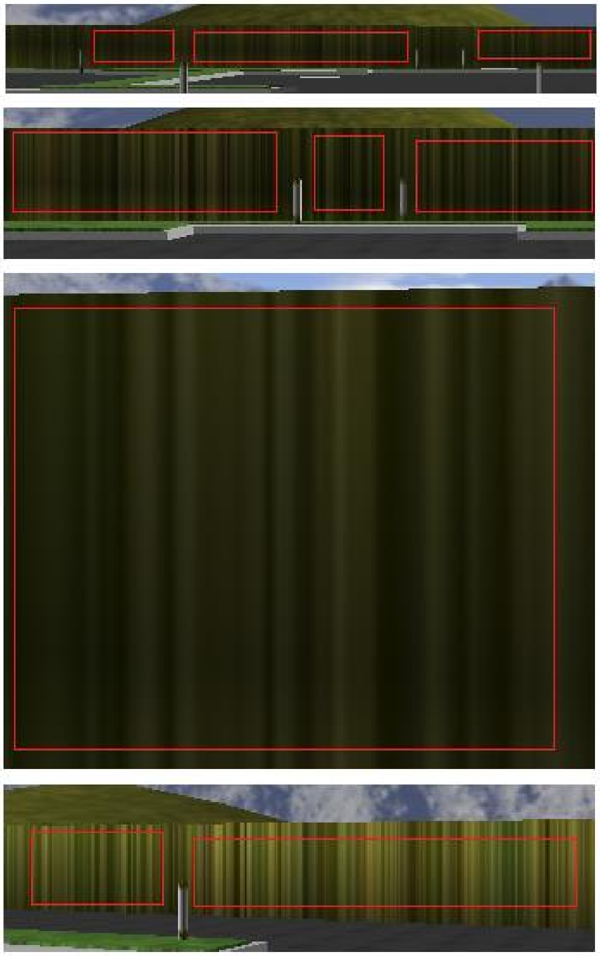

Fig. 4. Training data used for the calculation of μ and Σ.

the occupancy state. Since we do not have depth information from the images so all the cells along the line have high value of occupancy if the pixel is classified as belonging to object. The cells lying outside of the camera view have a probability value of 0.5 which means unknown state.

The steps to construct an occupancy grid for camera are as follows:

- Find pixels of the image corresponding to laser hit points for those laser beams which lie in the camera view. We need this step only to know the image region that is common with laser so that both occupancy grids are compatible.
- Find class of these pixels to see which of them belong to background and which of them belong to objects.
- Using definition of inverse sensor model given above, set probabilities of all cells using pixel classfication information.

Figure 5 shows an occupancy grid constructed from camera with laser hit points projected on this grid. We can see that most of the points lie in the area where there is high probability of occupancy, showing the effectiveness of this technique.

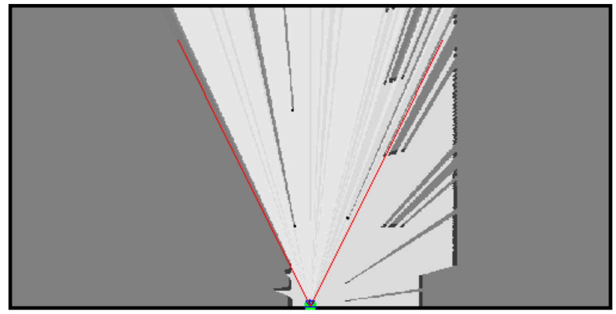

Fig. 5. Occupancy grid for camera within red lines with black dots representing objects detected by laser

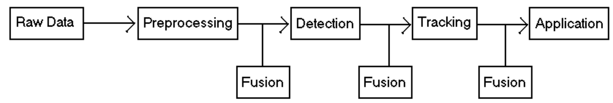

Fig. 6. Fusion at different levels

VI. Fusion

Fusion is the process of combining information from multiple sources in such a way that the combined information are more useful in some sense. Our objective is to combine information from laser and camera sensors. Fusion can be performed at different levels like, low level, object detection level and track level fusion as is shown in the figure 6. Our work is concerned with fusion at low level. An advantage of fusion at this level is that there is no information loss due to any interpretation of data which is necessary for higher levels.

There exist different fusion techniques of which we have used weighted linear combination of the two grids. Fusion equation is given below:

$$FusedGrid_{t,i} = \alpha.CameraGrid_{t,i} + (1-\alpha).LaserGrid_{t,i} \tag{8}$$

Here α is the weighting factor for camera and its value is between 0 and 1.

Another fusion technique that we have used is the conservative estimate. Let m_i^l is the i^{th} cell of the occupancy map built for laser sensor and m_i^c is the same cell for camera sensor. Then the value of same cell in the fused grid is given by:

$$m_i = max(m_i^l, m_i^c) \tag{9}$$

This map is the most pessimistic map given its components: If any of the sensor-specific map shows that a grid cell is occupied, so will the combined map. While this combined estimator is biased in factor of occupied maps, it is more appropriate than the Bayes filter approach when sensors with different characteristics are fused [11].

VII. Results

Fusion results of weighted linear combination for three different values of α are shown in figures 7, 8 and 9. The red lines in the fused view show the camera view with respect to laser view. The respective scenes for these results are shown

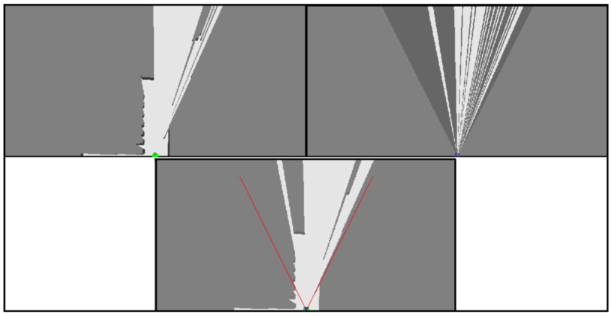

Fig. 7. Laser OG at left, Camera OG at right and Fused OG at bottom for $\alpha = 0.1$

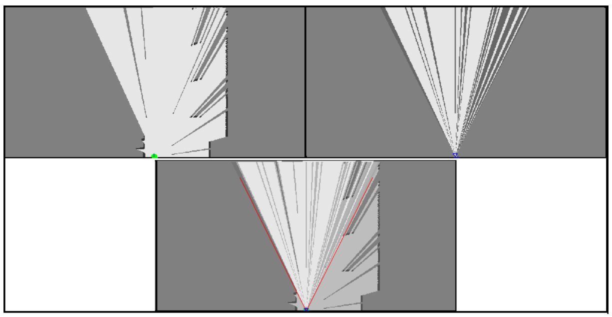

Fig. 8. Laser OG at left, Camera OG at right and Fused OG at bottom for $\alpha = 0.4$

in figures 10 and 11. Similarly the figure 12 shows fusion results for conservative approach for the same scene as shown in figure 11.

We can see from these results that the two occupancy grids constructed for laser and camera are very good approximation of the actual environment. And the fused view shows the combined information of both the sensors. Moreover the green rectangles in the scene figures show the detected objects.

VIII. Conclusion And Future Work

In this paper we have presented a technique to construct two occupancy grids one for laser and the other for camera. For camera we have introduced the idea of finding same region in the image that belongs to the scan plan of the laser sensor so

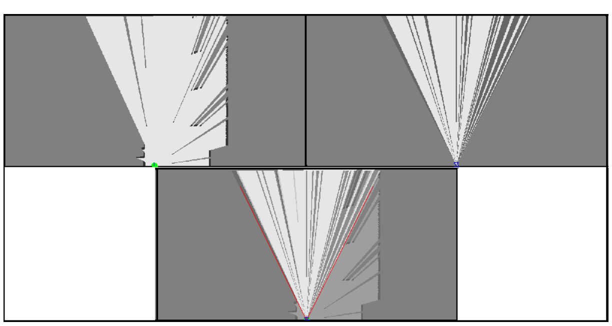

Fig. 9. Laser OG at left, Camera OG at right and Fused OG at bottom for $\alpha = 0.7$

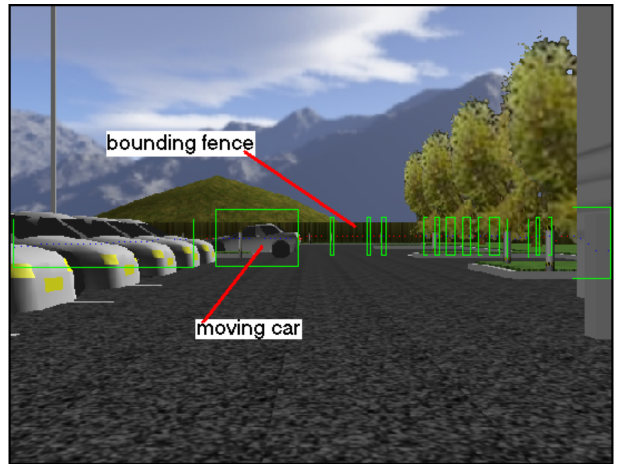

Fig. 10. Scene for $\alpha = 0.1$

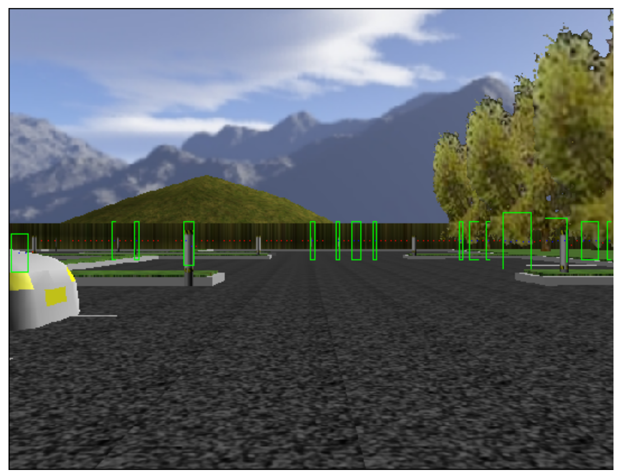

Fig. 11. Scene for $\alpha = 0.4$ and $\alpha = 0.7$

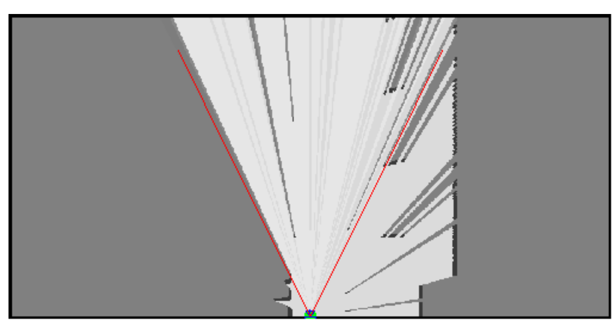

Fig. 12. Fusion result for conservative approach.

that the two constructed grids are compatible and can be further used for fusion. We have performed two kinds of low level fusion on these grids to get the combined view. The results show that this technique can be useful in the environments where background can be distinguished from the foreground.

For this work we had assumed that robot is not moving but objects in its view can be both dynamic and static. In future work we would apply these techniques with the robot moving and then to solve problems like Localization, SLAM, SLAMMOT AND DATMO. For camera we would also explore other techniques for object detection than background subtraction, this will enable us to apply these methods for robots functioning in unstructured environments. We believe that combining laser information we can be better able to detect regions of interest in camera images and this can be further used to get rich information of classification.

ACKNOWLEDGMENT

The authors would like to thank Higher Education Commission of Pakistan for supporting this work.

REFERENCES

[1] C. Blanc, L. Trassoudaine, Y. Le Guilloux, and R. Moreira. Track to track fusion method applied to road obstacle detection. In *International Conference on Information Fusion*, 2004.

[2] H. Cramer, U. Scheunert, and G. Wanielik. Multi sensor fusion for object detection using generalized feature models. In *International Conference on Information Fusion*, 2003.

[3] Alberto Elfes. *Occupancy grids: a probabilistic framework for robot perception and navigation*. PhD thesis, Carnegie Mellon University, 1989.

[4] J. Borenstein and Y. Koren. The vector field histogram - fast obstacle avoidance for mobile robots. *IEEE Journal of Robotics and Automation*, 7(3):278–288, June 1991.

[5] S. Thrun. Learning metric-topological maps for indoor mobile robot navigation. *Artificial Intelligence*, 1(99):21–71, 1999.

[6] D. Kortenkamp, R.P. Bonasso, and R. Murphy. *AI-based mobile robots*. MIT Press, 1998.

[7] TD. Vu, O. Aycard, and N. Appenrodt. Online localization and mapping with moving objects tracking in dynamic outdoor environments. In *IEEE International Conference on Intelligent Vehicles*, 2007.

[8] A. Elfes. *Occupancy grids: a probabilistic framework for robot percpetion and navigation*. PhD thesis, Carnegie Mellon University, 1989.

[9] D. Pagac, E.M. Nebot, and H. Durrant-Whyte. An evidential approach to map-building for autonomous vehicles. *IEEE Transactions on Robotics and Automation*, 14, 1998.

[10] G. Oriolo, G. Ulivi, and M. Vendittelli. Fuzzy maps: a new tool for mobile robot perception and planning. *J. of Robotic Systems*, 14(3):179–197, 1997.

[11] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. The MIT Press, September 2005.