Sensor Data Fusion Architecture for a Mobile Robot Application Using SIFT and Sonar Measurements

Alfredo Chávez Plascencia
Department of Electronic Systems
Automation and Control
Fredrik Bajers Vej 7C, 9220 Aalborg
Denmark
acp@es.aau.dk

Jan Dimon Bendtsen
Department of Electronic Systems
Automation and Control
Fredrik Bajers Vej 7C, 9220 Aalborg
Denmark
dimon@es.aau.dk

Abstract – *This paper presents and analyses the architecture of a process that can be used for map building and path planning purposes. This takes into account the uncertainty inherent in sensor measurements. To this end, Bayesian estimation and Dempster-Shafer evidential theory are used to fuse the sensory information and to update occupancy and evidential grid maps, respectively. The sensory information is obtained from a sonar array and a stereo vision system. Features are extracted using the Scale Invariant Feature Transform (SIFT) algorithm. A statistical comparison between both methods based on Mahalanobis distance measurement is carried out in the fused maps. Finally, the resulting two evidential maps based on Bayes and Dempster theories are used for path planning using the potential field method. The approach is illustrated using actual measurements from a laboratory robot. Both fusion techniques yield improved results, in comparison to using non-fused maps.*

Keywords: Sensor fusion, mobile robots, stereo vision, sonar, occupancy grids, SIFT, Dempster-Shafer, potential field.

1 Introduction

In the field of autonomous mobile robots one of the main requirements is to have the capacity to operate independently in uncertain and unknown environments; fusion of sensory information, map building and path planning are some of the key capabilities that the mobile robot has to possess in order to achieve autonomy. Map building must be performed based on data from sensors; the data in turn must be interpreted and fused by means of sensor models. The fusion process can be carried out using various data fusion methods [2]. The result of the fusion of the sensor information is utilized to construct a map of the robot's environment and the robot can then plan its own path, avoiding obstacles along the way.

The sensor fusion data algorithms considered in the scope of this paper are: Bayesian method, Dempster-Shafer method, Fuzzy Logic and Artificial Neural Networks, [2, 22, 1, 23]. Each sensor fusion method previously mentioned is unique to some extend. The Bayesian is the oldest approach and the one with strongest foundation. The Dempster-Shafer method is a recent attempt to allow more interpretation of what uncertainty is all about. Both methods offer approaches to some of the fundamental problems of sensor fusion: information uncertainty, conflicts, and incompleteness [24]. Due to this fact, the inclination of using Bayes and Dempster-Shafer approaches have been taken into consideration to carry out the fusion process along the research in this paper.

This paper extends the work done in [21], where a sensor data fusion approach to map building is presented. The approach is exemplified by building a map for a laboratory robot by fusing range readings from a sonar array with landmarks extracted from stereo vision images using the SIFT algorithm. The paper also shows that it is feasible to perform path planning based on the potential field derived from maps that have been generated using fused range readings from the sonar and the vision system.

In this paper, an architecture for a sensor data fusion application to map building is proposed. It also contributes with the comparison of two sensor fusion techniques: Bayesian Inference and Dempster-Shafer Evidential theory. The comparison is carried out based on the Mahalanobis distance method.

These techniques also yield so-called Occupancy and Dempster-Shafer grids, respectively, which are internal map representations that can be used for robot navigation. Occupancy grids were introduced by Elfes in [3, 4]. Dempster-Shafer grids were proposed in [5], as an alternative to occupancy grids. Localisation can also be implemented, but it is not considered in this paper.

The paper is organised as follows. A multilayer hierarchical structure for a sensor data fusion is addressed in section 2. An overview of sensor models and sensor fusion is presented in sections 3 and 4, along with the main contribution of this paper: a novel sensor fusion of Scale Invariant Feature Transform, a recently developed computer vision

method [11], and sonar range readings. Section 5 outlines how the sensor fusion can be employed to generate potential field for indoor robot path planning upon which, experimental results are presented in section 6 where experiments are based on Bayes and Dempster theories, a comparison between both sensor fusion techniques (Bayes and Dempster) are showed as well as path planning experiments based Potential field. Finally, the conclusion of this work is summed up in section 7.

2 Hierarchical Structure

In this section a multilayer hierarchical structure (architecture) for an application in multisensor data fusion in mobile robots is presented. Figure 1 illustrates how the structure is divided into layers which is based on the functional diagram on multisensor integration and fusion shown in [1, 2].

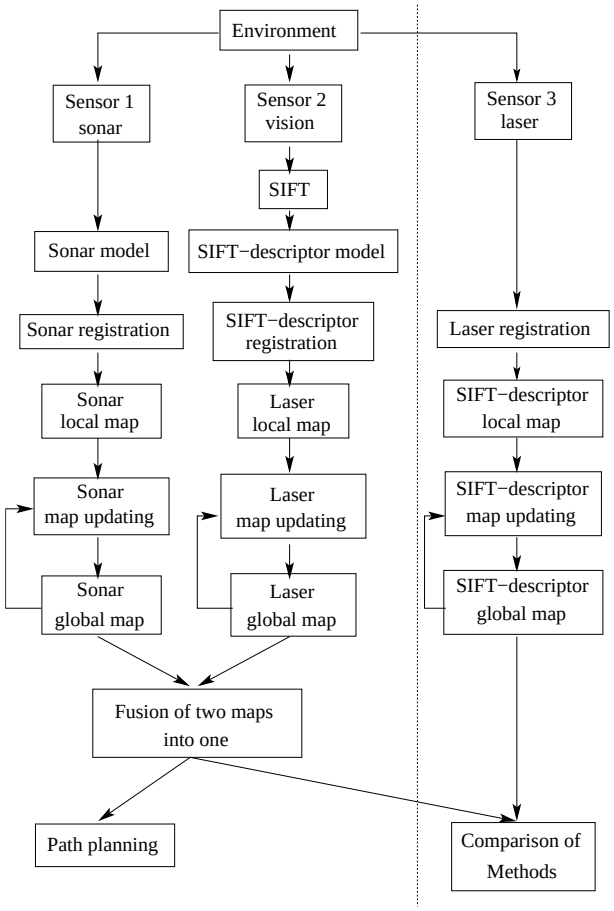


Figure 1: Architecture of the map building process.

- $Layer_1$ represents the environment (laboratory) in which the robot has to navigate and collect data.
- $Layer_2$ presents the sensors used for the task and their data acquisition. The mobile robot uses sensors to interact with its environment, the following of which have been considered in this application are: an ultrasonic sound device or sonar (it measures the distance to objects). A stereo vision system (it takes pairs of snapshots from the scene during the travel of the mobile robot). A laser range finder is used for evaluating the incoming data from sonar and vision system, but not used for the fusion or map building processes.
- $Layer_3$ contains the SIFT algorithm which extracts features from digital images.
- $Layer_4$ depicts the sensor data models which are necessary to quantify the uncertainty that always comes with the sensor data.
- $Layer_5$ considers the registration of the sensor data models into a $2D$ spatial representation which is common to all sensors. This means that sensor readings of different modalities must be converted into common internal representations in advance before the fusion process is carried out. Occupancy and Dempster-Shafer grids are the two such spatial representations used in this work.
- $Layer_6$ shows a local map for each type of sensor which represents a window of the environment around the robot. The window is normally centered at the robot's current position.
- $Layer_7$ addresses the sonar map updating. Every time the mobile robot gathers data while travelling along its path, it computes a new local map ($Layer_6$) that can be used to construct a global map ($Layer_8$). Bayes and Dempster-Shafer update rules are used in both sonar-based and stereo-base mapping systems.
- $Layer_8$ deals with a global map which is simply an abstraction of the entire environment the mobile robot has been in. This map can then be used to plan an optimal path for the robot from its initial state to a goal state.
- $Layer_9$ shows the integration of data provided by two qualitative different sensors; the sensor data fusion is carried out by integrating the two maps into one [14].
- $Layer_{10}$ concerns itself with the path planning approach based on the potential field derived from maps that have been generated using fused range readings from the sonar and vision systems. Sensor fused data from both fused modalities (Bayes and Dempster-Shafer) are compared with each other. The comparison is done by applying the Mahalanobis distance to the maps.

3 Sensor Models

3.1 Sonar Model

A common sensor used to measure distance is the ultrasonic range finder, a.k.a. sonar. The sonar can measure the

distance from the transducer to an object quite accurately. However, it can not estimate at what angle within the sonar cone the pulse was reflected. Hence, there will be some uncertainty about the angle at which the obstacle was measured. A wide range of sonar models have been developed in the past years by various researchers, [3], [4], [12], and [15]. Taking the starting point in these methods, a grid G of cells $C_{i,j}, 1 \leq i = x \leq x_{max}, 1 \leq j = y \leq y_{max}$ is defined in front of the sensor.

Consider the representation of the sonar beam cone shown in figure 2, where the sonar beam is formulated as two probability density functions. These functions measure the confidence and uncertainty of an empty and occupied region in the cone beam of the sonar respectively. They are defined based on the geometrical aspect and the spatial sensitivity of the sonar beam.

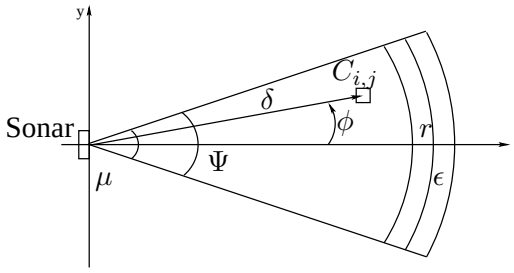


Figure 2: Sonar Model

Let Ψ denote the top angle of the cone in the horizontal plane and let ϕ denote the (unknown) angle from the centerline of the beam to the grid cell $C_{i,j}$. Let r denote a sonar range measurement and ϵ the mean sonar deviation error. The value μ in the sonar model represents the minimal measurement and δ is the distance from the sonar to the cell. Using μ, the sonar model can combine quadratic and exponential distributions in the empty probability region of the sonar model [15]. Then $P_s^e(\delta, \phi, r) = F_s(\delta, r)A_n(\phi)$ represents the probability of the cell $C_{i,j}$ (translated from polar coordinates (r, ϕ)) being empty, and $P_s^o(\delta, \phi, r) = \mathcal{O}_s(\delta, r)A_n(\phi)$ represents the probability of the cell $C_{i,j}$ being occupied. The factors F_s, $\mathcal{O}_s$ and $A_n(\phi)$ are given by equations 1, 2 and 3, [15].

$$F_s(\delta, r) = \begin{cases} 1 - \left(\frac{\delta}{r}\right)^2, & \text{if } \delta \in [0, \mu] \\ e^{\delta}, & \text{if } \delta \in [\mu, r - \epsilon] \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$\mathcal{O}_s(\delta, r) = \begin{cases} \left(\frac{1}{r}\right)\left(1 - \left(\frac{\delta - r}{\epsilon}\right)^2\right), & \text{if } \delta \in [r - \epsilon, r + \epsilon] \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

and

$$A_n(\phi) = \begin{cases} 1 - \left(\frac{2\phi}{\Psi}\right)^2, & \text{if } \phi \in \left[-\frac{\Psi}{2}, \frac{\Psi}{2}\right] \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

3.2 Vision-SIFT-descriptor Model

3.2.1 SIFT

The other sensor used for sensor fusion in this study is a stereo vision system. In particular, the Scale Invariant Feature Transform (SIFT) is a method for extracting distinctive invariant features from digital images [11]. The features are invariant to scaling and rotation. They also provide a robust matching across a substantial range of affine distortion, change in $3D$ view point, addition of noise and change in illumination. Furthermore, the features are distinctive, i.e. they can be matched with high probability to other features in a large database with many images. The SIFT algorithm consists of the following major steps:

- Scale-space peak detection: the aim of this step is to find locations in the image that are invariant to scale change in the same image.
- Accurate key-point localization: in this step the position of each point candidate is determined; points with low contrast and poor localization along the edge are removed. This yields a so-called descriptor at each point.
- Majority orientation assignment: this step makes the rotation descriptor invariant. This is done by assigning a consistent orientation to each key-point.
- Computation of the local image descriptor: this step associates each feature point with a 128-element feature vector or interest point descriptor that uniquely identifies that point.

Once the descriptors are found in each image, i.e. left and right images, a matching algorithm is applied in both images. Figure 3 presents the matching features descriptors which have been identified from a stereo pair of images.

Figure 3: Descriptor matches between left and right images.

3.3 SIFT-descriptor model

3.3.1 Camera Model

The camera model used is the so-called perspective or pinhole model. This model can be seen in the left part of the figure 4. It represents the camera by its optical center C_l, an image plane π, the camera frame, the focal length f, the optical axes and the principal point (O_x, O_y). A $3D$

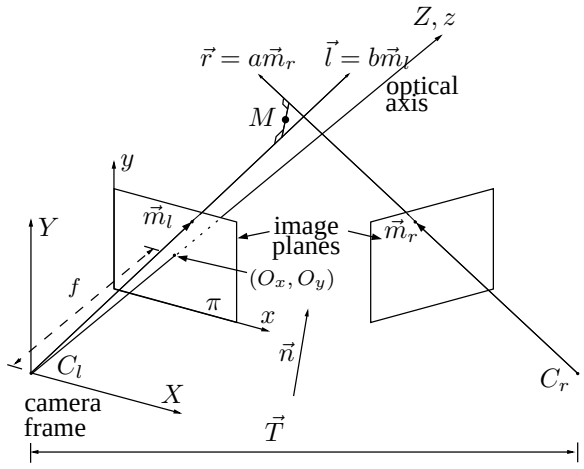


Figure 4: The figure shows in the left part the perspective or pinhole model and also the triangulation.

vector $M = [X, Y, Z]^T$ in a real world coordinate frame which point $M = (X, Y, Z)$ is projected into the image plane or camera frame as the vector $\vec{m}_l = [x, y, z]^T$.

3.3.2 Triangulation

The triangulation algorithm outlined in [20] has been implemented in order to obtain the depth of the matching SIFT descriptors.The problem of this algorithm is to determine the midpoint of the vector of the segment parallel to the vector $\vec{n}$ that joins $\vec{l} = am_l$, and $\vec{r} = \vec{T} + bR^T\vec{m}_r$, where $a \in \mathbb{R}$, and $b \in \mathbb{R}$, as it is shown in figure 4. The $3D$ coordinates of each descriptor can be obtained by solving the equation (4), and the coefficients a, b, c can be obtained by solving the linear system (5). The Z component in equation (4) represents the depth to the point M.

$$\vec{M} = [X, Y, Z]^T = a\vec{m}_l + \frac{1}{2}c(\vec{m}_r \times \vec{m}_r) \qquad (4)$$

$$a\vec{m}_l - bR^T + c(\vec{m}_l \times R^T\vec{m}_r) = \vec{T} \qquad (5)$$

R and $\vec{T}$ are defined as the intrinsic and extrinsic parameters of the stereo system, where, $R = R_l R_r^T$ and $\vec{T} = \vec{T}_r - R^T\vec{T}_l$. $(\vec{T}_l,\ R_l)$ and $(\vec{T}_r,\ R_r)$ are the intrinsic and extrinsic parameters of the two cameras in the world reference frame. These parameters have been obtained from the Matlab Camera Calibration Toolbox using single camera calibration. $\vec{m}_l$ and $\vec{m}_r$ are the projection vectors of the $3D$ point into the left and right image planes respectively.

3.3.3 Stereo Triangulation Error

Due to the factors of quantification and calibration errors, a certain degree of uncertainty must be expected in the triangulation. Mathies and Shafer [13] shows how to model and calculate the triangulation error in stereo matched with $3D$ normal distributions. Geometrically these uncertainties translate into ellipsoidal regions. The stereo uncertainty error and the $3D$ Gaussian distribution can be depicted as in figures 5(a) and 5(b).

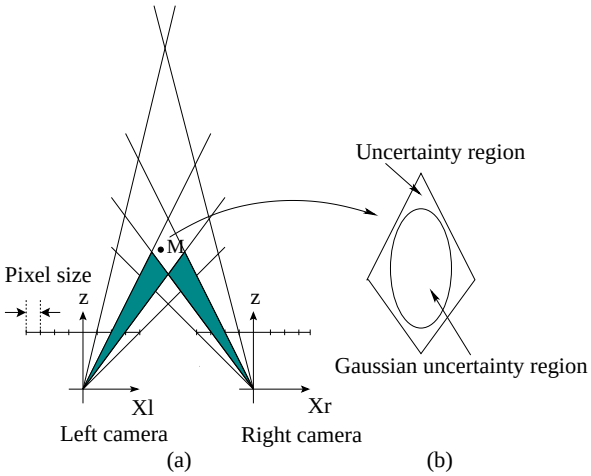


Figure 5: ((a) Stereo Geometry showing triangulation uncertainty as a diamond around a point M. It also shows the empty region uncertainty from the pair of cameras to the uncertainty region of the point M. (b) $2D$ dimensional Gaussian distribution uncertainty region.

In the following, the details of the triangulation error modelling are shown, [13]. Consider a $3D$ point $M = (X, Y, Z)$, which is projected onto the left and right image planes respectively as $\vec{m}_l = [x_l, y_l]$ and $\vec{m}_r = [x_r, y_r]$ as they can be depicted in figure 4. These vectors are assumed to be normally distributed with means μ_l and μ_r and covariance matrices V_l and V_r. The covariance matrix V_p of the point M is shown in equation 6, [13].

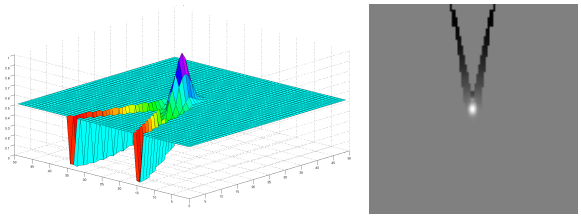
$$V_p = J \begin{bmatrix} V_l & 0 \\ 0 & V_r \end{bmatrix} J^T \qquad (6)$$

Where J is the jacobian of first partial derivatives of $f(m_l, m_r) = M$. These equations can be obtained by solving the equation (4).

The empty regions from the left and right cameras as shown as shadow areas in figure 5(a) also need to be modelled. The approach taken by Elfes in [3] to model the empty region of the sonar beam has been taken into consideration and implemented with satisfactory results which can be depicted as in figures 6(a) and 6(b). Figure 6(a) shows a $3D$ probability SIFT-descriptor model. Figure 6(b) shows a $2D$ view of the $3D$ model.

4 Sensor Fusion

In the following, two different sensor fusion techniques are applied: 1) Bayesian theory and 2) Dempster-Shafer theory of evidence.



(a) (b)

Figure 6: (a) $3D$ view of the SIFT-descriptor probability model. (b) $2D$ view of the SIFT-descriptor probability model.

4.1 Bayes Theory

4.1.1 Bayes Update Formula

Elfes and Matthies [3] have proposed in their previous work the use of a recursive Bayes formula to update the occupancy grid for multiple sensor observations $(r_1, ..., r_t, r_{t+1})$. When this formula is transferred to the occupancy grid framework, the following is obtained:

$$P^o = \frac{P_s^o P_m^o}{P_s^o P_m^o + (1 - P_s^o)(1 - P_m^o)} \tag{7}$$

- P_m^o and $1 - P_m^o$ are the prior probabilities that a cell is occupied and empty respectively; they are taken from the existing map.
- P_s^o is the conditional probability that a sensor would return the sensor reading given the state of the cell being occupied. This conditional probability is given by the probabilistic sensor model $(P_s^o(\delta, \phi, r))$.
- P^o is the conditional probability that a cell is occupied based on the past sensor readings. It is the new estimate.

A new sensor reading, introduces additional information about the state of the cell $C_{i,j}$. This information is done by the sensor model P_s^o and it is combined with the most recent probability estimate stored in the cell. This combination is done by the recursive Bayes' rule based on the current set of readings $r_t = (r_t, r_{t-n}, ..., r_0)$ to give a new estimate P^o. It is worth noting that when initializing the map an equal probability to each cell $C_{i,j}$ must be assigned. In other words, the initial map cell prior probabilities are $P_m^o = 1 - P_m^o = 0.5 \; \forall C_{i,j}$.

4.1.2 Fusion of Sensors With Two Occupancy Grids

In this method, an occupancy grid based on the Bayes' rule is constructed for each sensor type, which will then be fused to build up the resulting grid map. Afterwards, the cells in each grid map are modified in order to reinforce the cell probability of being occupied, [14]:

- The probability of a cell $C_{i,j}$ being occupied is set to one if it is higher than a predefined threshold T_o.
- The probability of a cell $C_{i,j}$ being occupied is reinforce if it is between the interval $[\frac{1}{2}, T_o]$. The reinforcement strength the probability of a cell being occupied.
- Otherwise the value in the cell $C_{i,j}$ remains.

More precisely, the resulting grid map is computed in two steps.

Firstly, probability values in the grid maps are modified for each sensor type using the following expression:

$$P_{n=1,2}^o(c) = \begin{cases} 1 & \text{for } P^o(c) > T_o, \\ \frac{P^o(c) + T_o - 1}{2 \cdot T_o - 1} & \text{for } P^o(c) \in [\frac{1}{2}, T_o] \\ P^o(c) & \text{otherwise} \end{cases} \tag{8}$$

Where $P^o(c)$ is the probability of the cell $C_{i,j}$ being occupied. $P_1^o(c)$ is the modified probability of occupancy from the first sensor and $P_2^o(c)$ is the modified probability of occupancy from the second sensor.

Secondly, the computed values are then inserted in Bayes' rule to obtain the occupied fused probability $P_f^o(c)$ of the cell $C_{i,j}$ in the resulting grid.

$$P_f^o(c) = \frac{P_1^o(c) P_2^o(c)}{P_1^o(c) P_2^o(c) + (1 - P_1^o(c)(1 - P_2^o(c))} \tag{9}$$

4.2 Dempster-Shafer Theory

The second method concerns Dempster-Shafer theory of evidence. This theory was proposed by Glenn Shafer [5] as an extension of the work presented in [6] and [7].

Dempster-Shafer theory is mainly characterized by a frame of discernment $(FOD = \Theta)$, a basic probability assignment function (bpa), a belief function (Bel) and a plausibility (PLS) function. These are tied together via the so-called Dempster's rule of combination [8].

Each proposition in Θ is called a singleton. 2^Θ is called the power set of Θ. Any subset of Θ is called a hypothesis. Applying the notion of frames of discernment to an occupancy grid yields a set of frames $\Theta_{i,j} = \{o, e\}$; where i, j represents an individual cell in the grid. Let A denote the subsets of the power set of $2^{\Theta_{i,j}} = 2^{\{o,e\}} = \{\{\varnothing\}, \{o\}, \{e\}, \{o, e\}\}$; where $\{\varnothing\}$ and $\{o, e\}$ are the empty and the $disjunction$ or $dontknow$ subsets, respectively. $\{o\}$ and $\{e\}$ denote the probabilities of the cell being occupied or empty, respectively. The *quantum of belief* is distributed as $Bel(A) = m(\varnothing) + m(o) + m(e) + m(o, e) = 1$, [5]. Finally, the function $m : \; 2^\Theta \rightarrow [0, 1]$ is called the basic probability assignment. and must satisfy the following criteria.

$$\sum_{A \subset 2^\Theta} m(A) = 1 \tag{10}$$

$$m(\varnothing) = 0 \tag{11}$$

Equation (11) reflects the fact that no belief is assigned to $\varnothing$. In order to obtain the total evidence assigned to A, one must

add to $m(A)$ the quantities $m(B)$ for all proper subsets B of A.

$$Bel(A) = \sum_{\forall B:B\subseteq A} m(B) \tag{12}$$

In [5], the notion of *plausibility* or upper probability of A is defined as $1 - Bel(\neg A)$; where $(\neg A)$ is used to denote the set theoretic complement of A. $Bel(\neg A)$ is the disbelief of the hypothesis of A. Consequently, $Pls(A)$ can be thought of as the amount of evidence that does not support its negation. All in all, this sums up to

$$Pls(A) = 1 - Bel(\neg A) = 1 - \sum_{\forall B:B\not\subset A} m(B) \tag{13}$$

Notice that $Bel(A) \leq Pls(A)$ for any given A.

The above assumptions brings up with a formulation of a formal Dempster's rule of combination.

Suppose Bel_1 and Bel_2 are belief functions over the same frame of discernment Θ, with basic probability assignment in each focal element, e.g. $\{m(A_1), \cdots, m(A_k)\}$ and $\{m(B_1), \cdots, m(B_l)\}$ respectively. The belief function given by $m(C_k)$ is called orthogonal sum of Bel_1 and $Bel2$ and is represented as $Bel_1 \oplus Bel2$,

$$m(C_k) = \frac{\sum_{\forall A_i,B_j\in 2^{\Theta_{i,j}}:A_i\cap B_i=C_k;C_k\neq\emptyset} m(A_i)m(B_j)}{1 - \sum_{\forall A_i,B_j\in 2^{\Theta_{i,j}}:A_i\cap B_j=\emptyset} m(A_i)m(B_j)} \tag{14}$$

Some remarks are drawn. These two belief functions are independent and have at least one focal element in common. The two belief functions can be combined by finding the focal intersections for each C_k, where C is the set of all subsets produced by $A_i \cap B_j$. The denominator in equation 14 is the normalisation term.

When using Dempster's rule of combination to update a grid map for each cell $C_{i,j}$ lying in the main lobe of the sonar model and for each interpreted sensor reading, equation (14) becomes:

$$m^o = \frac{m_o^G m_o^S + m_o^G m_{o,e}^S + m_{o,e}^G m_o^S}{1 - m_e^G m_o^S - m_o^G m_e^S} \tag{15}$$

$$m^e = \frac{m_e^G m_e^S + m_e^G m_{o,e}^S + m_{o,e}^G m_e^S}{1 - m_e^G m_o^S - m_o^G m_e^S} \tag{16}$$

$$m^{o,e} = \frac{m_{o,e}^G m_{o,e}^S}{1 - m_e^G m_o^S - m_o^G m_e^S} \tag{17}$$

The quantities m_o^S, m_e^S and $m_{o,e}^S$ are obtained from sensor models, while m_o^G, m_e^G and $m_{o,e}^G$ are obtained from the existing grid map. Note that $m_{o,e}^G = 1 - m_o^G - m_e^G$, and $m_{o,e}^S = 1 - m_o^S - m_e^S$. m^o, m^o, and $m^{o,e}$ are the new updates. All cells $C_{i,j}$ in the Shafer grid map are initialized as stated in (18) since there is no *apriori* knowledge of evidence.

$$\left.\begin{aligned} m_o^G &= 0 \\ m_e^G &= 0 \\ m_{o,e}^G &= 1 \end{aligned}\right\} \quad \forall C_{i,j} \in G \tag{18}$$

The above assumption means total ignorance about the state of each cell. However, when the mobile robot is moving and gathering data from the environment it uses this data to update the map using Dempster's rule of combination equations 15, 16, and 17.

The lack of ignorance is depicted in the expression (19), and simply expresses an exact knowledge of the environment.

$$\left.\begin{aligned} m_o^G + m_e^G &= 1 \\ m_{o,e}^G &= 0 \end{aligned}\right\} \quad \forall C_{i,j} \in G \tag{19}$$

5 Path Planning Using Potential Field

The main idea of potential field is to discretize the configuration space $\mathcal{W}$ of the robot $\mathcal{A}$ into a regular grid and search for an appropriate path τ within that grid. In this approach, the robot is considered as a particle in the configuration space moving under the influence of an artificial potential field U. The potential field consists of the sum of an attractive potential field generated by the goal and a repulsive potential generated by the obstacles [18], as seen in equation (20).

$$U(\vec{q}) = U_{att}(\vec{q}) + U_{rep}(\vec{q}) \tag{20}$$

where $\vec{q} = [x, y, \theta]^T$ (a compact set closed and bounded in $\mathcal{W}$) which is the current state of the robot (a.k.a. *configuration*). An example of an attractive and a repulsive potential field functions can be depicted in equations 21 and 22.

$$U_{att}(\vec{q}) = \frac{1}{2}\xi\rho_{goal}^2(\vec{q}) \tag{21}$$

$$U_{rep}(\vec{q}) = \begin{cases} \frac{1}{2}\eta\left(\frac{1}{\rho(\vec{q})} - \frac{1}{\rho_0}\right)^2 & \text{if } \rho(\vec{q}) \leq \rho_0, \\ 0 & \text{if } \rho(\vec{q}) > \rho_0, \end{cases} \tag{22}$$

Where ξ and η are a positive scaling factors. ρ_{goal} denotes the Euclidean distance between the current and the goal configurations, i.e. $\|\vec{q} - \vec{q}_{goal}\|$. $\rho_{(\vec{q})}$ is the Euclidean distance from the current configuration of the robot $\vec{q}$ to the obstacle region $\mathcal{CB}$. ρ_0 is the maximum distance of influence, i.e. it is the distance from the center of the obstacle to the boundary of the obstacle region.

The force to attract and repulse the robot can be obtained from the negated gradient of the potential.

$$\vec{F} = -\nabla U(\vec{q}) = -\begin{bmatrix} \frac{\partial U(\vec{q})}{\partial x} \\ \frac{\partial U(\vec{q})}{\partial y} \end{bmatrix}$$
$$= -\begin{bmatrix} \frac{\partial U_{att}(\vec{q})}{\partial x} + \frac{\partial U_{rep}(\vec{q})}{\partial x} \\ \frac{\partial U_{att}(\vec{q})}{\partial y} + \frac{\partial U_{rep}(\vec{q})}{\partial y} \end{bmatrix}$$
$$= -\begin{bmatrix} \frac{\partial U_{att}(\vec{q})}{\partial x} \\ \frac{\partial U_{att}(\vec{q})}{\partial y} \end{bmatrix} - \begin{bmatrix} \frac{\partial U_{rep}(\vec{q})}{\partial x} \\ \frac{\partial U_{rep}(\vec{q})}{\partial y} \end{bmatrix}$$
$$= \vec{F}_{att}(\vec{q}) + \vec{F}_{rep}(\vec{q}) \quad (23)$$

The potential field can be obtained mathematically when the position of the obstacles are precisely identified. The obstacles generate a repulsive potential field which makes the robot navigate far from the obstacles. The other option considered in this article consists of moving the robot through the obstacles generated by, applying the sensor fusion techniques (Bayes and Dempster's rules) to the sensor readings. The attractive potential field is added to the potential field generated from the environment using sensor readings.

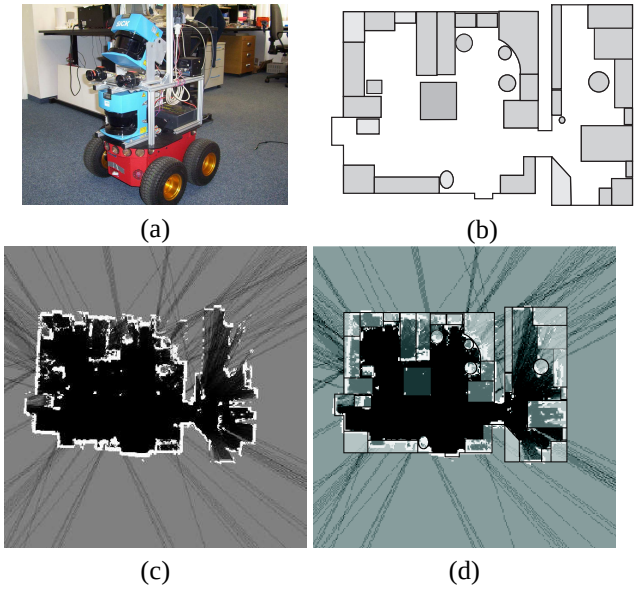


Figure 7: (a) Pioneer3AT from ActiveMedia Robotics. (b) Layout of the laboratory/office. (c) Represents the map of the laboratory/office based laser readings. (d) The map of the laboratory/office embedded into the laser map.

6 Experimental Results

A Pioneer3AT from ActiveMedia Robotics, as shown in figure 7(a), serves as an experimental testbed. It provides data by using a ring of 16 ultrasonic sensors, a stereo vision system and a laser rangefinder. The laser rangefinder was used for the purpose of evaluating the incoming data from the sonar and the stereo pair of cameras respectively. The experiment was carried out in a typical laboratory environment, the layout of which can be depicted in figure 7(b). Figure 7(c) shows the grid created only from laser rangefinder data. This picture demonstrates the shape of the room. The layout of the laboratory is embedded into the laser map as, seen in figure 7(d). It can be seen that the laser map is quite accurate when compared with the layout of the laboratory. For this reason it is utilized as a reference map.

The experiments consist of gathering 30 measurements during the robots' motion. A sonar array measurement consists of gathering 16 sonar readings. A vision measurement consists of a single snapshot. Finally, a laser measurement consists of gathering 360 readings.

6.1 Experiments based on Bayes Theory

This section presents the experiment results of map making based on SIFT-features descriptors, sonar readings and the fusion over the sonar and SIFT-descriptors maps using the recursive Bayes' formula.

Matching descriptors were found in each stereo pair of images by the SIFT-algorithm. Figure 8(a) represents the result of the process in which the SIFT-descriptors matching each stereo image were fused and plotted. This process was made along all the stereo snapshots taken by the robot on its path. This map is sparse due to the number of snapshots taken by the robot during the experiment that were not numerous enough to construct a denser map of the laboratory. However, the map presents very important information about narrow free spaces that can be seen in figure 8(d). The narrow space is situated between the two rooms and detected correctly. It can also be seen in this figure that there are some SIFT-features which have been plotted outside the layout of the laboratory. The reason being that there is a natural limitation in the implemented technique of the SIFT-algorithm where some descriptors are not matched correctly causing a misreading, consequently producing a bad plot or in other words a plot outside of the laboratory map. In can also be observed that many features where detected over the desks. This is due to the stereo vision system being placed at different level than the sonar ring as seen in figure 7(a).

The whole picture of the sonar data fusion process can be depicted in figure 8(b) which shows the grid created only from sonar data and in figure 8(e) a layout of the laboratory is embedded in the sonar map. It can be noticed that the space before and inside the door, i.e. the space between the two rooms is poorly detected. Usually, the sonar system has a problem with the detection of doors and similarly shape narrow places. This phenomenon is created by the shape of the sonar beam. It can also be observed that the number of detected objects over the desks is smaller than the SIFT-map.

Figure 8(c) shows the result of applying the method stated in 4.1.2. It can be seen that the empty as well as the occupied areas are reinforced when they are compared with the two individual maps. Figure 8(f) shows the grid embedded into the office's map.

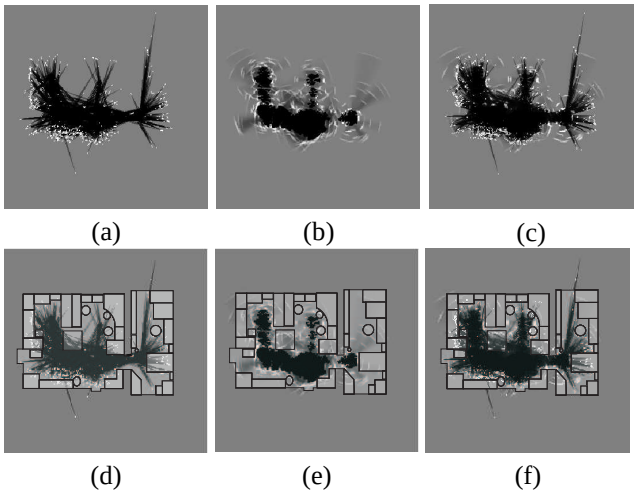


Figure 8: Maps generated from applying Bayes theory to interpreted sensor readings. Top row: (a) SIFT-descriptor map. (b) sonar map. (c) SIFT-sonar fused maps ; Bottom row: maps with office layout superimposed.

6.2 Experiments based on Dempster-Shafer Theory

The experiments depicted in figure 9, show the result of applying formulas 15 and 16 to the interpreted sensor range readings. The interpretation is done using probabilistic sensor models as the ones described in section 3.

Figure 9(a) shows the $vision-occupied$ grid, which represents the evidence of being occupied for the vision system. The white dots represent an evidence of an area being occupied. The black color beyond the white dots is a region where the mobile robot does not know anything about it, i.e. there is a total ignorance. The black area in the middle of the plot and which is surrounding by the white dots represent lack of ignorance and a low evidence of occupation is assigned to this region.

Figure 9(d) represents the $sonar-occupied$ grid, which represents the evidence of occupation by the sonar ring system. The white arcs represent the evidence of an occupied area. The black area inside the arcs or in the middle of the grid represents the empty area or lack of ignorance, meaning that $zero$ evidence of occupation is assigned to this area. The gray color situated between the white arcs and the empty region is the transition between the occupied and empty regions respectively. The black area beyond the white arcs represent total ignorance.

Figure 9(g) shows the $sonar-vision-occupied$ grid, which is the result of fusing the $vision-occupied$ grid map from the vision system (figure 9(a)) with the $sonar-occupied$ grid from the sonar ring system (figure 9(d)). By analyzing the figure; it can easily be seen that the evidence of an occupied area has increased (white color). Uncertainty in the occupied region of the sonar beam has been reduced. The empty area (black area in the middle of the grid) has been reinforced in comparison to the individual

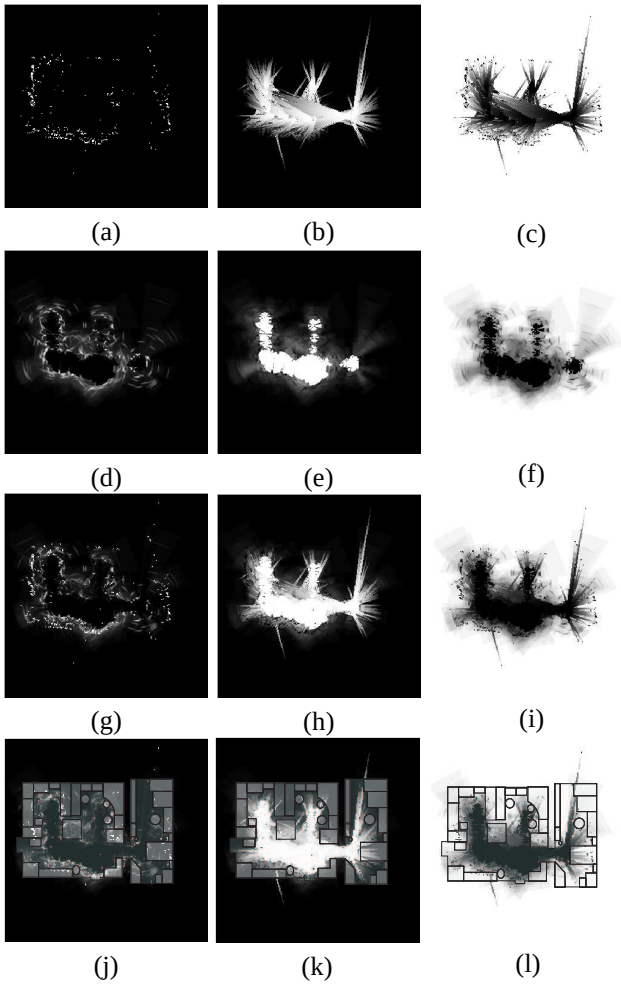


Figure 9: Maps generated from m^o, m^e and $m^{o,e}$. (a) $vision-occupied$ map. (b) $vision-empty$ map. (c) $vision-dontknow$ map. (d) $sonar-occupied$ map. (e) $sonar-empty$ map. (f) $sonar-dontknow$ map. (g) $vision-sonar-occupied$ map. (h) $vision-sonar-empty$ map.(i) $vision-sonar-dontknow$ map. Bottom row; $vision-sonar$ maps with office layout superimposed.

areas from vision and sonar grid maps respectively. The state of the black area which surrounds the white area is unknown.

Figure 9(b) represents the $vision-empty$ grid, which represents the evidence of being empty for the vision system and empty evidence is explored. At first glance, two color areas can be distinguished, black and white. The white area is the lack of ignorance, i.e. high empty evidence can be assigned to the cells. But, looking carefully at this area, it can be seen that the white color changes smoothly in comparison to the black i.e. gray color, which represents low evidence of emptiness. The black area beyond the white area signifies total ignorance of evidence and since the dealing is with the $vision-empty$ grid the terms m_o^G and $m_{o,e}^G$ are discarded from expression 18 in the plotting of this area and just the term m_e^G is taken into account which gets the value

of 0.

Figure 9(e) shows the $sonar - empty$ grid, which represents the evidence of being empty for the sonar ring system. The white area represents the empty region. The gray area is the transition between the empty and occupied regions. The black area beyond the white area signifies a total ignorance of evidence.

Figure 9(h) shows the result of fusing the resulting $vision - empty$ grid map from the vision system (figure 9(b) with the resulting $sonar - empty$ grid map from the sonar ring system (figure 9(e)). Looking carefully at the figure; it can easily be noticed that the evidence of an empty area (white color in the middle of the plot) has increased, i.e the empty area has been reinforced. It can be seen when it is compared with the individual areas from $vision - empty$ and $sonar - empty$ grid maps respectively. The black area surrounding the white area is the total ignorance. There is a gray area between the black and white areas which is the transition of emptiness (white) to the total of ignorance (black). There are some black zones within the gray area, which represent strong evidence of occupation.

Figure 9(c) shows the $vision - dontknow$ grid which, represents the evidence of disjunction for the vision system. The black area in the middle of the map signifies lack of ignorance. Although, one can see that the black color changes smoothly from black (middle of the map) to gray and then to black (dot pots) and then to white. Representing ignorance of evidence.

Figure 9(f) shows the result of applying equation 17 to the interpreted sonar data which generates the $sonar - dontknow$ grid. The meaning of the colors are explained in the following. The black color means lack of ignorance and high evidence can be assigned to the empty area. The gray color is the level of transition from lack of ignorance to total ignorance, meaning that the empty evidence goes from being high to low. The dark arcs inside the cones of the sonar beam represent strong evidence of occupation. The white area beyond the cones of the sonar beam represents total ignorance of evidence.

Figure 9(i) shows the $sonar - vision - dontknow$ grid, which is the result of fusing the resulting $vision - dontknow$ grid map from the vision system (figure9(c)) with the resulting $sonar - dontknow$ grid map from the sonar ring system (figure 9(f)). The black area in the middle of the plot signifies lack of ignorance. Thus a high degree of empty evidence can be assigned to that area. The gray area is the transition from the empty area to the occupied area or in other words, it is the transition from lack of ignorance to total ignorance. During the transition, gray arcs and black dots can be seen. The arcs are the occupied region of the sonar beam; the more black the arcs are the more the evidence of the arcs being occupied. The black dots are the SIFT-features, which reinforce the occupied region of the sonar beam. The white surface means total ignorance of evidence.

Figures 9(j), 9(k) and 9(l) show the $vision - sonar - occupied$, $vision - sonar - empty$ and $vision - sonar - dontknow$ grid maps embedded into laboratory map respectively.

6.3 Mahalanobis Distance Comparison

The Mahalanobis distance measure approach was introduced by [25] in 1936. It is based on correlations between random vectors. It differs from Euclidean distance in that it takes into account the correlations of the data set.

Lets $\vec{x}$ and $\vec{y}$ be two random vectors, the Mahalanobis distance d_M from a vector $\vec{y}$ to the vector $\vec{x}$ is the distance from $\vec{y}$ to $\hat{\vec{x}}$, the centroid of $\vec{x}$, weighted according to $\mathbf{C_x}$, the covariance matrix of $\vec{x}$, so that,

$$d_M = ((\vec{y} - \hat{\vec{x}})' {\mathbf{C_x}}^{-1} (\vec{y} - \hat{\vec{x}}))^{\frac{1}{2}} \quad (24)$$

$$Where:$$

$$\hat{\vec{x}} = \frac{1}{2} \sum_{i=1}^{n_x} \vec{x_i} \quad (25)$$

$$\mathbf{C_x} = \frac{1}{n_x - 1} \sum_{1}^{n_x} (\vec{x_i} - \hat{\vec{x}})(\vec{x_i} - \hat{\vec{x}})' \quad (26)$$

The Mahalanobis distance from a SIFT, sonar, and, SIFT-sonar vectors to a laser, is computed in the following. The elements of the SIFT, sonar, and SIFT-sonar vectors are the coordinates of the occupied cells of their respective maps. The elements of the laser vector are also the coordinates of the occupied cells of its respective map. The laser is taken as a true parameter vector to be compared with the other vectors.

The Mahalanobis distance is computed in squared units of each observation in the reference sample $\vec{x}$. A unit has a value of $5\ cm$ which is the size of a single cell in the grid.

A 2D grid plot (laser & SIFT), which has been generated by the laser grid map and the SIFT grid map based on Bayes approach, is presented in figure 10(a). The red squares correspond to the occupied laser cells. The asterisks represent the occupied cells by the $SITF$-descriptor grid map. Each colour represents a Mahalanobis distance to the laser vector. The corresponding colour values of the distances are represented as a colour bar placed next to the map. Figure 10(c) depicts the plot of the Mahalabobis distance from figure 10(a). The same situation for the Dempster approach is depicted in figures 10(b) and 10(d). A comparison of these two plots reveals that both SIFT-descriptor grids based on Bayes and Dempster approaches approximate the laser plot. The difference stems from the fact that the SIFT-feature algorithm finds features in the scene that the laser is not able to find and vice versa, the laser & SIFT (Dempster) distance plot is significantly less abundant than the laser & SIFT (Bayes) distance plot.

The situation where the sonar coordinates vector is taken into account to compute the Mahalanobis distance to a laser coordinates vector can be depicted in figure 11. The colour of the asterisks in figures 11(a) and 11(b) are yellow, blue,

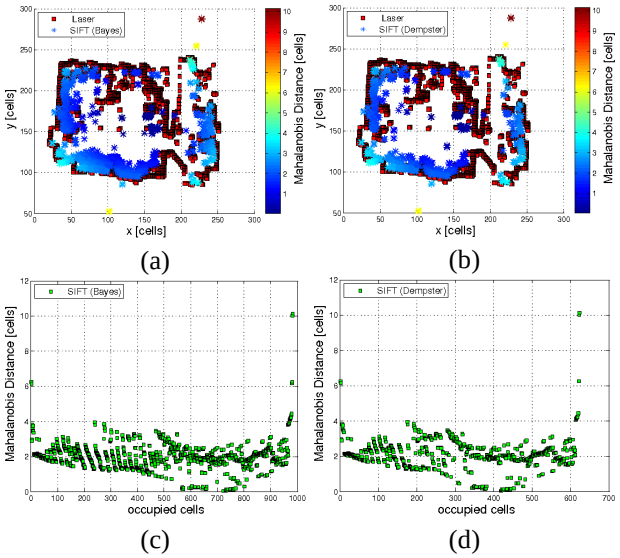


Figure 10: (a) Mahalanobis distance between the SIFT-descriptor and Laser maps based on Bayes. (b) Shows only the Mahalanobis distance from (a). (c) Mahalanobis distance between the SIFT-descriptor and Laser maps based on Dempster. (d) Shows only the Mahalanobis distance from (b).

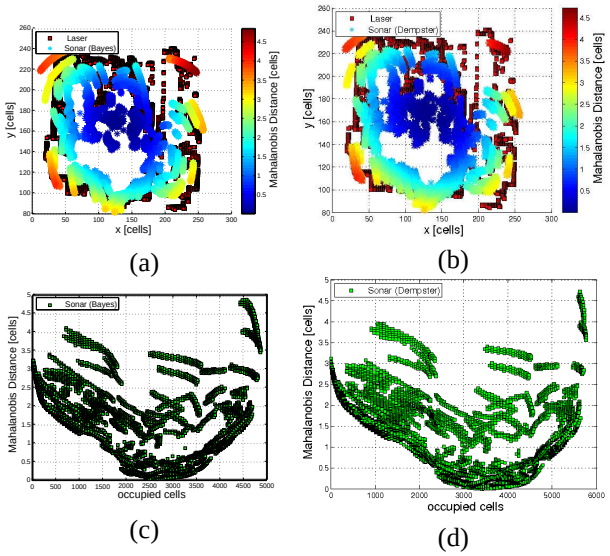


Figure 11: (a) Mahalanobis distance between the sonar and laser maps based on Bayes. (b) Shows only the Mahalanobis distance from (a). (c) Mahalanobis distance between the sonar and laser maps based on Dempster. (d) Shows only the Mahalanobis distance from (b).

and red, meaning that the Mahalanobis distance mainly oscillates from $\frac{1}{2}$ to 4.0 square units where few cells are outside of this range. This situation is more clear in figures 11(c) and 11(d) where the concentration of cells is under 4 square units.

The representation of the occupied cells of the fusion between the sonar and the SIFT-descriptor maps is presented in figure 12. The occupied cells are represented by the asterisks, and their colours are mainly blue. This means that the Mahalanobis distance from the SIFT-sonar coordinates vector to the laser coordinates vector mainly oscillates between $\frac{1}{2}$ to 3.5 square units.

Table 1 presents the number of occupied cells, the mean and variance values of the maps that have been analysed in this subsection. The number of cells in both $vision-sonar$ based on Bayes and $vision-sonar$ based on Dempster approaches have been reduced when comparing with individual sonar sensor maps. This can be attributed to the fact that, when fusing the sonar map with the SIFT-descriptor map, many inaccurate cells are cancelled. Another point to notice is that the mean value concerning the $vision-sonar$ map based on Dempster's approach is less than the one of Bayes. This can be interpreted as the $vision-sonar$ map based on Dempster approach be more accurate to the true laser map. Both $vision-sonar$ based on Dempster and Bayes variances disperse about its mean with almost the same value. Further comparison of table 1 shows that Dempster-Shafer performs better in terms of definition of an occupied area.

The success of the Dempster-Shafer method is attributed to the following characteristics of this method [26]:

- Dempster-Shafer theory allows support for more than one proposition at a time, rather than a single hypothesis as Bayes does.
- The uncertainty interval bounded by Pls and Bel allows the lack of data (ignorance) to be modelled adequately.
- Dempster-Shafer theory does not require prior probabilities to function, However it does require some preliminary assignment of masses that reflects the initial knowledge of the system [27].

Table 1: Summarises the number of the occupied cells, the mean and the variance values of the grid maps.

Bayes			
Grid Map	Occupied cells	mean	Variance
$vision$ (SIFT)	983	1.9921	0.6948
$sonar$	4855	1.6692	1.0746
$vision-sonar$	4272	1.8558	1.1155
Dempster			
$vision$ (SIFT)	624	2.0990	0.8715
$sonar$	5775	1.50995	0.9356
$vision-sonar$	5712	1.7991	1.1184

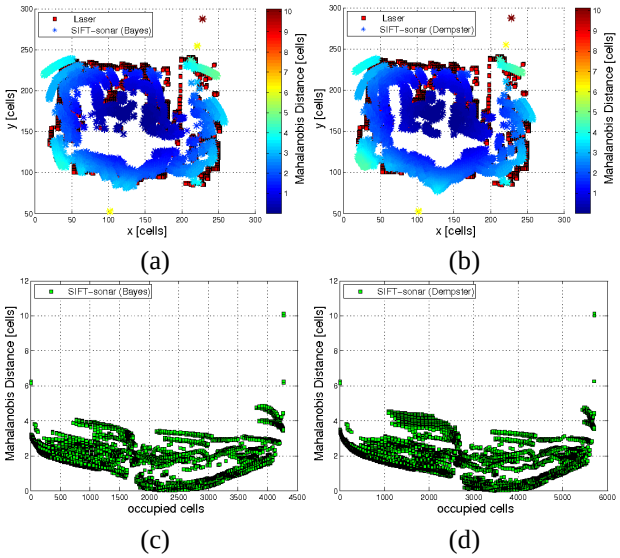


Figure 12: (a) Mahalanobis distance between the SIFT-sonar and laser maps based on Bayes. (b) Shows only the Mahalanobis distance from (a). (c) Mahalanobis distance between the SIFT-sonar and laser maps based on Dempster. (d) Shows only the Mahalanobis distance from (b).

6.4 Path planning Experiments based on Potential Field

The algorithm implemented in this section is called "Depth-first planning" [18]; it mainly consists of constructing single segments starting at the initial configuration of the robot $\vec{q}_{init}$. The direction of each segment is obtained by solving equation 23; this technique simply follows the steepest descent of the potential function until the goal configuration $\vec{q}_{goal}$ is reached. A drawback of this method is that the mobile robot may get trapped into a local minimum, which did not occur in the present simulation. However, solutions to the local minima problem exist [19].

Figure 13 shows an example of Depth first path planning algorithm whereby the potential field approach has been applied to the maps from figures 8(c), 9(g), 9(h) and 9(i).According to this, figure 13(a) corresponds to the $vision - sonar$ (Bayes) map (figure 8(c)). Figure 13(b) corresponds to the $vision - sonar - occupied$ (Dempster) map (figure 9(g)). Figure 13(c) corresponds to the $vision - sonar - empty$ (Dempster) map (figure 9(h)). Figure 13(d) corresponds to the $vision - sonar - dontknow$ (Dempster) map (figure 9(i)). The shape of the laboratory environment is embedded into these grid maps where a path can be seen which connects a start point configuration with a final goal configuration. The area which surrounds the path represents the empty region which is bounded by the occupied area.

7 Conclusion and Future Work

This article presents a multilayer hierarchical structure (architecture) for multisensor data fusion for robot map making and path planning. The work considers the use of Bayes and Dempster-Shafer rules of combination to integrate sensor readings from a stereo vision system using the SIFT algorithm and a ring of sonars. The experiments were verified with real data in a real indoor environment. The experiments show that the use of the SIFT algorithm can improve the sonar map and it can be effectively used for robot path planning. When comparing the two fusion techniques, Dempster Shafer has better definition of the occupied area than Bayes. On the other hand, the Mahalanobis distance measure shows that the Dempster approach is more accurate to the true laser map than the Bayesian approach. Future research work is to apply control strategies to follow the path planned by the algorithm.

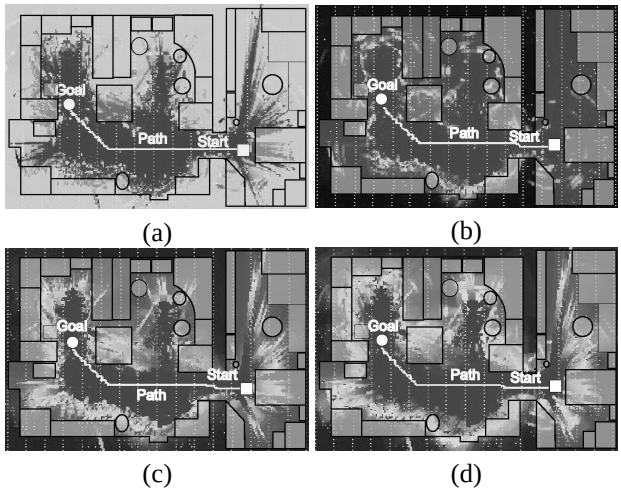


Figure 13: Figures (a), (b), (c) and (d) depict a path planned by the algorithm "Depth-first planning".

References

[1] Ren C. Luo and Michael G. Kay, "A Tutorial on Multisensor Integration and Fusion", *Proceedings of the IEEE Conference of Industrial Electronics Society, IECON '90., 16th Annual Conference of IEEE*, Vol. 1, pp. 707–722, 1990.

[2] R.C. Luo and C.C. Yih and K.L. Su, "Multisensor Fusion and Integration: Approaches, Applications, and Future Research Directions", *IEEE Sensors Journal*, Vol. 2, No. 2, 2002.

[3] P. Moravec and A. Elfes, "High Resolution Maps from Wide Angle Sonar", *IEEE Journal of Robotics and Automation*, pp. 116–121, 1985.

[4] A. Elfes, "Using Occupancy Grids for Mobile Robot Perception and Navigation", *IEEE Journal of Robotics and Automation*, Vol. 22, pp. 45–57,1989.

[5] G. Shafer, "A Mathematical Theory of Evidence", *Princeton, NJ:Princeton University. Press*, 1976.

[6] Dempster A. P., "A generalization of Bayesian Inference", *Jornal of the Royal Statistical Society*, Vol. 30, pp. 205–247, 1968.

[7] Dempster A. P., "Upper and Lower probabilities Included by a Multivalued Mapping", *Annals of Mathematical statistics*, Vol. 38, pp. 325–339, 1967.

[8] Daniel Pagnac, Eduardo M. Nebot and Hugh Durrant-Whyte, "An Evidential Approach to Map Building for Autonomous Vehicles", *IEEE Transactions of Robotics and Automation*, Vol. 14, pp. 623–629, 1998.

[9] Alfredo Chávez and Jan Dimon Bendtsen, "Sensor Fusion -Sonar and Stereo Vision, Using Occupancy Grids and SIFT", *Proceedings of the IEEE International Conference on Dynamics, Instrumentation and Control*, pp. 303–314, 2006.

[10] Alfredo Chávez and Petr Štěpán, "Sensor Data Fusion", *Proceedings of the IEEE Systems, Man and Cybernetics Society, Conference on Advances in Cybernetics Systems*, pp. 20–25, 2006.

[11] David G. Lowe, "Distinctive Image Features from Scale-Invariant Key points", *International Journal of Computer Vision* Vol 2, pp. 91–110, 2004.

[12] Kurt Konolige, "Improved Occupancy Grids for Map Building", *Autonomous Robots* Vol 4, pp. 351–367, 1997.

[13] Larry Matthies and Steven A. Shafer, "Error Modeling in Stereo Navigation", *IEEE Journal of Robotics and Automation*, Vol 3, pp. 239–248, 1987.

[14] Štěpán,P., Miroslav Kulich and Libor Přeučil, "Robust Data Fusion with Occupancy Grids", *IEEE Transactions on Systems, Man, and Cybernetics* , Vol. 35, pp. 106-115, 2005.

[15] Štěpán,P. Libor Přeučil, and Luboš Král "Statistical Approach to Integration Interpretation of Robot Sensor Data ", *In Proceedings of the IEEE International Workshop on Expert Systems Applications.*, pp. 742-747, 1997.

[16] Petr Štěpán "Vnitřní Reprezentance Prostředí Pro Autonomní Mobilní Roboty (User Interface for Autonomous Mobile Robots)", *PhD thesis.*, 2001.

[17] Andrew Howard, Maja J Matarić and Gaurav S Sukhatme "Mobile Sensor Network Deployment using Potential Fields: A Distributed, Scalable Solution to the Area Coverage Problem", *Proceedings of the 6th International Symposium on DARS02, Fukuoka, Japan*, 2002.

[18] Jean Claude Latombe "Robot Motion Planning", *Stanford University*, 1991.

[19] Christopher I. Connolly and J. B. Burns Sukhatme "Path Planning Using Laplace's Equation", *In Proceedings of the IEEE International Conference on Robotics and Automation*, pp 2102–2106, 1990.

[20] Emanuele Trucco and Alessandro Verri "Introductory techniques for 3-D Computer Vision, Prentice Hall", 1998.

[21] A. Chávez and H. Raposo "Robot path Planning Using SIFT and Sonar Sensor Fusion", *Proceedings of the 7th WSEAS International Conference on Robotics, control and Manufacturing Technology, ROCOM'07, Hagzhou, China*, pp 251–256, 2007.

[22] R.C. Luo and M.G. Kay "Multisensor Integration and Fusion in Intelligent Systems", *IEEE Transactions on Systems, Man, and Cybernetics*,Vol. 19, No. 5, pp 901–931, 1989.

[23] J.K. Hackett and M. Shah "Proceedings of the IEEE International Conference on robotics and Automation, 1990", *Proceedings of the IEEE International Conference on robotics and Automation*,Vol. 2, pp 1324–1330, 1990.

[24] J. Braun "Dempster-Shafer theory and Bayesian reasoning in multisensor data fusion, Sensor Fusion: Architectures, Algorithms and Applications.", *Procedings of SPIE*, pp 255–266, 2000.

[25] P.C. Mahalanobis "On the generalized distance in statistics", *Proceedings of the National Institute of Science of India*,Vol. 12, pp 49–55, 1936.

[26] D. Pagac and E.M. Nebot and H. F. Durrant-Whyte "An Evidential Approach to Map Bulding for Autonomous Vehicles", *IEEE Transactions of Robotics and Automation*,Vol. 14, No. 4, pp 623–629, 1998.

[27] S. Challa and D. Koks "Bayesian and Dempster-Shafer fusion", *Sadhana*,Vol. 29, pp 145–174, 2004.

Alfredo Chávez Plascencia received the M.Sc. in Intelligent Autonomous Systems and PhD in Sensor Fusion in mobile robots from Aalborg University, Denmark, in 2002 and 2008 respectively. He current research interests include nonlinear control, artificial intelligence and vision withing the field of mobile robots.

Jan Dimon Bendtsen received the M.Sc. and PhD in control engineering from Aalborg University, Denmark, in 1996 and 1999, respectively. He is currently employed at Aalborg University as Associate Professor. In 2005, he was visiting researcher at Australian National University in Canberra. He was co-receiver of the Best Technical Paper award at the AIAA Guidance, Navigation, and Control Conference 2006. His current research interests include identification and control of uncertain nonlinear systems, in particular systems with time-varying structure.