Robot Navigation Map Building Using Stereo Vision Based 3D Occupancy Grid

Haythem Ghazouani, Moncef Tagina, René Zapata

Robot Navigation Map Building Using Stereo Vision Based 3D Occupancy Grid

H. Ghazouani *, **M. Tagina** **, **R. Zapata** *

* *LIRMM Laboratory,Department of Robotic, University of Montpellier II, 161 rue Ada, 34392 Montpellier Cedex 5, France*
tel: +33 636 904 947
e-mails: haythemghz@yahoo.fr, zapata@lirmm.fr
** *SOIE Laboratory, Department of Applied Computer Science, National School for Computer Studies, Campus Universitaire de La Manouba 2010, La Manouba, Tunisia*
e-mail: moncef.tagina@ensi.rnu.tn

Submitted: 12/07/2010
Accepted: 24/09/2010
Appeared: 15/11/2010

Abstract: In this paper environment is modeled using depth information provided by a stereo vision system. Workspace is decomposed into voxels which are the smallest volume of environment. A first observation on the state of the voxels is calculated based on stereo system provided 3D points and triangulation error propagation. A new method for model update using prior and current observations on the voxel state is presented. The proposed update function uses a credibility value that denotes how strongly a new observation shall influence the voxel state based on the age of the last observation and the homogeneity of the current observations. Finally, the 3D occupancy grid is scaled down to a 2D map to reduce computational costs. Experimental results using real environment and comparison based on a benchmarking method are presented to demonstrate the performance of our approach.

Keywords: Stereo Vision, 3D Occupancy Grid, Map Building.

1. INTRODUCTION

Robot navigation can be performed using a map of the environment. In the case of an unknown environment the robot must build its own representation of it. Robotic mapping has been an active area in artificial intelligence for a few decades. It addresses the problem of acquiring a spatial model of the workspace through available sensors on a robot. Information from sensors is processed and model of the environment is updated. The characteristics of a good map representation must be able to quickly update its knowledge about the current state of the environment without heavy computation effort. At any iteration of map building the measurements will have a slight inaccuracy, and then any features being added to the map will contain corresponding errors. If unchecked, these errors build cumulatively grossly distorting map. One of the greatest difficulties of map building arises from the nature of the inaccuracies and uncertainties in terms of noise in sensor measurements, which often lead to inaccurate maps. In this paper, input information comes from a depth map produced by a stereo vision system. Disparity values are converted into real distances using triangulation and a 3D occupancy grid is constructed incrementally based on the positions of 3D points and the defined 3D occupancies (voxels). The construction of the incremental 3D model takes into account the propagation of camera calibration error and matching error. For the update of the 3D model, a credibility value based on the homogeneity of the observations on a local neighborhood and the age of the last prior observation is proposed.

The paper is divided as follows. The next section gives a survey of occupancy grid based map building methods. Section 3 gives an overview over the whole map building system. Section 4 introduces our method for map building. The main contribution of this section is the modeling of triangulation error, the use of a new update function for the 3D grid states and the 2D map cell state discretization. Section 5 presents experimental results using Pioneer 3 equipped with two cameras and comparison with other paradigms using Collins et al. benchmarking suite (Collins et al. 2007).

2. SURVEY OF OCCUPANCY GRID BASED MAP BUILDING METHODS

The occupancy grids, also known as evidence grids or certainty grids were pioneered by Moravec and Elfes (Moravec & Elfes, 1985; Elfes, 1987, Moravec, 1988; Elfes, 1989a; Elfes, 1989b) and formulated in the Carnegie Mellon University (Martin & Moravec, 1996) as a way to construct an internal representation of static environments by evenly spaced grids based on ultrasonic range measurements. Occupancy grids provide a data structure that allows

fusion of sensor data. It provides a representation of the world which is created with inputs from the sensors. Apart from being used directly for sensor fusion, there also exist interesting variations of evidence grids, such as place-centric grids (Youngblood, 2000), histogram grids (Koren & Borenstein, 1991) and response grids (Howard & Kitchen, 1996). Occupancy Grids is certainly the state of the art method in the field of grid based mapping. It is the most widely used robot mapping technique due to its simplicity and robustness and also because it is flexible enough to accommodate many kinds of spatial sensors with different modalities and combining different sensor scans. It also adapts well to dynamic environments.

In general, the occupancy grid technique divides the environment into two dimensional discrete grid cells. In a stochastic occupancy grid (Badino et al., 2007) the intensity of each cell denotes the likelihood that a world point is at the lateral position and depth represented by the cell. A world point can therefore occupy more than one cell. How large an area the world point affects depends on the variance (noise) associated with the point. There exists various occupancy grid representations.

The occupancy grids map is considered as a discrete state stochastic process defined over a set of continuous spatial coordinates. Each grid cell is an element and represents an area of the environment. The state variable associated with any grid cell C_i in the grid map yields the occupancy probability value of the corresponding region. Since the probabilities are identified based on the sensor data, they are purely conditional. Given a sensor data, each cell in the occupancy grid can be generally in two states $s(C_i) = Occupied$ or $s(C_i) = Free$, and to each cell there is probability $P[s(C_i) = Occupied]$ attached, which reflects the belief of the cell C_i being occupied by an object.

$$P[s(C_i) = Free] = 1 - P[s(C_i) = Occupied] \qquad (1)$$

Occupancy grids have been implemented with laser range finders (Schmid et al. 2010) stereo vision sensors (Moravec, 1996) and even with a combination of sonar, infrared sensors and sensory data obtained from stereo vision (Lanthier et al., 2004). Recent works of occupancy grid map building have focused on stereo vision as input sensor. Franco and Boyer presented a method for visual occupancy grid using multi camera environment (Franco & Boyer, 2005). The idea of their method is to consider each camera pixel as statical occupancy sensor. All pixel observations are then used jointly to infer where, and how likely, matter is presented in the scene. Kenji et al. attempt to eliminate false positive in a stereo vision obstacle detection method. For this purpose, they propose a method that generates Occupancy Grid Maps based on measurements from a stereo vision system which leads to robust obstacle detection (Kohara et al. 2010). Braillon et al. have proposed a real-time method to detect obstacles using theoretical models of the ground plane, first in a 3D point cloud given by a stereo camera, and then in an optical flow field given by one of the stereo pairs' camera (Braillon et al. 2006). The idea of their method is to combine two partial occupancy grids from both sensor modalities with an occupancy grid framework. In (Oniga et al., 2009), the authors have used an occupancy grid computed with a method that outputs an occupancy grid with three distinct cell types: road, traffic isles and obstacles. They have performed a temporal filtering of the false traffic isles present in the grids. Obstacle cells were separated into static (probably infrastructure) and dynamic. An enhanced occupancy grid was built, containing road, traffic isle, static obstacle and dynamic obstacle cells. The global map was obtained by integrating the enhanced occupancy grid along several successive frames. Lategahn et al. present in (Lategahn et al. 2010) a complete processing chain for computing 2D occupancy grids from image sequences. First the 3D points reconstructed from the images are distributed onto the underlying grid. Thereafter a virtual measurement is computed for each cell thus reducing computational complexity and rejecting potential outliers. Subsequently a height profile is updated from which the current measurement is partitioned into ground and obstacle pixels.

In (Lu et al. 2010), the authors gives the different techniques for building occupancy grid map. These techniques are based on Bayesian theory (probabilistic approach) (Moravec, 2001) (Elfes, 1992), Dempster Shafer theory of evidence (evidence theoretic approach) (Ribo & Pinz, 2001)(Gambino et al. 1996), and fuzzy set theory (possibility approach) (Oriolo et al. 1999) (Ribo & Pinz, 2001)(Gambino et al. 1996).

The use of probability theory in occupancy grids based approach has been criticized for several reasons. Firstly, it is difficult to create accurate sensor models for new sensors. The characteristics of ultrasonic sensors are well known, but unrealistic simplifications are needed to model the complex behavior of stereo vision. Hence, some authors even decided to skip probability theory and to invent an own update rule (Guadarrama & Ruiz-Mayor, 2010). Secondly, a single probability does not allow to distinguish between *unknown* and *uncertain* occupancy. Thus, it can not be determined whether an area has not been scanned at all (e.g. due to occlusion) or the sensor data was unreliable. For these reasons, we present in this work a new approach to determine and update the states of the occupancies.

3. OVERVIEW OVER THE WHOLE MAP BUILDING SYSTEM

Figure 1 gives an overview of the map building system. The input are disparity images and the parameters of the stereo camera. This paper concentrates on creating a visual map from stereo and motion data using occupancy grid approach.

A stereo vision system we have developed in (Ghazouani et al. 2010) is used to deliver dense disparity maps. An adaptive support window is used to have more reliable disparities. We use a camera motion estimation method proposed by Hirschmüller et al. (Hirschmüller et al. 2002) that utilizes the knowledge about the three dimensional position of features (i.e. corners, which are detected in the left image) to robustly and accurately determine motion of the camera. Motion is calculated between two consecutive stereo images without any pre-knowledge or prediction about feature location or the possibly large camera movement.

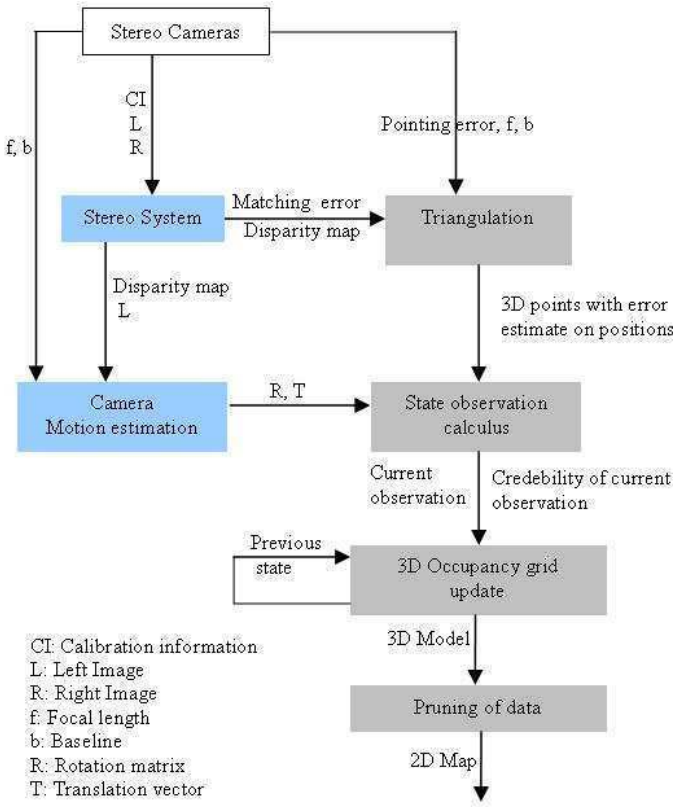

Fig. 1. Overview over the whole system.

First, triangulation is used to calculate the positions of 3D points from the provided stereo data. Then, a method to determine the state of the cells is proposed based on the triangulation error estimation. A new update function of the cells is presented. Finally the 3D grid is scaled down to a 2D map for optimization reasons. The proposed map building process (gray squares) is presented in section 4.

4. MAP BUILDING SYSTEM

In this section map building steps are described. We begin by calculating the triangulation error propagation. Then, we use a 3D occupancy grid to model the environment based on the calculated error and the provided stereo data. After that, we propose an update function for the environment model. Finally, we generate a 2D navigation map based on the projection of the 3D model.

4.1 Triangulation error propagation

The stereo vision system constructs a 3D view of the environment and then extracts the obstacles present in it. This is done by either generating disparity maps and evaluating them (Zhang et al. 2009; Boyer et al. 1991), or generating 3D occupancy grid of the environment using the disparity values. 3D occupancy grid can provide good reconstruction quality at low cost. The resulting accuracy of the three-dimensional position information provided by the stereo vision process is a crucial point for quality control tasks.

Over the last years, some efforts have been spent on error analysis in stereo-vision based computer vision systems (Blostein & Huang, 1987)(Yang & Wang, 1996)(Ramakrishna & Vaidvanathan, 1997)(Kamberova & Bajcsy, 1998)(Balasuramanian et al. 2000)(Rivera-Rios et al. 2005)(Park & Subbarao, 2005)(Albouy et al. 2006). As an example, Blostein and Huang (Blostein & Huang, 1987) have investigated the accuracy in obtaining 3D positional information based on triangulation using point correspondences derived using a stereoscopic camera setup. They have derived closed form expressions for the probability distribution of position errors along each direction (horizontal, vertical and range) of the coordinate system of the stereo rig. Also, a study of different types of error and their effects on 3D reconstruction results obtained using a structured light technique has been presented by Yang and Wang (Yang & Wang, 1996). In their work, Yang and Wang have derived expressions for the errors observed for the 3D surface position, the orientation and the curvature measurements. Further, Ramakrishna et al. (Ramakrishna & Vaidvanathan, 1997) proposed a new approach for estimating tight bounds on measurement errors, considering the inaccuracies introduced during calibration and triangulation. Balasuramanian et al.(Balasuramanian et al. 2000) analyzed the effect of noise (which is assumed to be independent and uniformly distributed) and of the geometry of the imaging setup on the reconstruction error for a straight line, their analysis being mainly based on simulation studies. Revira-Rios et al. (Rivera-Rios et al. 2005) have analyzed the error when measuring dimensionally line entities, these errors being mostly due to localization errors in the image planes of the stereo setup. Consequently, in order to determine optimal camera poses, a non-linear program has been formulated, that minimizes the total MSE (Mean Square Error) for the line to be measured, while satisfying sensor related constraints. Lastly, the accuracy of 3D reconstructions has been evaluated through comparison with ground truth in contributions presented by Park et al. (Park & Subbarao, 2005) and Albouy et al. (Albouy et al. 2006) More recently, Jianxi et al.(Jianxi et al. 2008) have presented an error analysis for 3D reconstruction taking into account only the camera calibration parameter accuracy.

In a stereo vision system based on binocular cameras, each 3D point P projects onto the left image at $U_l = [u_l, v_l]$ and onto the right image at $U_r = [u_r, v_r]$. Because of errors in measurement, the stereo system will determine U_l and U_r with some error, which in turn causes error in the estimate location of the real point P. We want to take this uncertainty into account in any reasoning based on measurements of P, specially in the calculation of grid occupancy state.

Once we have found conjugate pairs in the two images, it is possible to get the depth of the correspondent points in the scene if we know: the mutual position of the cameras (extrinsic parameters) and the sensors parameters (intrinsic parameters). We define a 3D disparity space which dimensions are u, v and d respectively to designate row, column and disparity. Each element (u, v, d) of the disparity space is projected to pixel (u, v) in the reference image and the pixel $(u, v' = v + d)$ in the matching image. Each disparity pixel in the stereo image can be converted to a 3D point based on the projective camera equations. For our system, with the cameras aligned so that the

optical axes are parallel these equations are given by the following equations:

$$\begin{cases} x = \dfrac{zu}{f} \\ y = \dfrac{zv}{f} \\ z = \dfrac{fb}{d} \end{cases} \quad (2)$$

where $U = (u, v)$ is the position of the disparity pixel in the reference camera image plane, $X = (x, y, z)$ is the position of the observed 3D point in the reference camera coordinate frame and d is the pixel disparity. We define our sensor model to be made of two parts: pointing error: p and matching error m. Pointing error is the error in the position of the vector U of the reference camera and is based on the accuracy of the camera calibration. Matching error is the accuracy of the disparity, d, and is based on the accuracy of the correlation algorithm. These accuracies are features of the stereo camera. Given these values, the covariance matrix of the disparity pixel in (u, v, d) space is given by the following equation :

$$C_U = \begin{bmatrix} p & 0 & 0 \\ 0 & p & 0 \\ 0 & 0 & m \end{bmatrix} \quad (3)$$

To obtain the covariance matrix C_X of the 3D point (x, y, z) associated with a disparity pixel (u, v, d), we propagate the error from (u, v, d) space to (x, y, z) space by applying the methods given in Faugeras (Faugeras, 1997). The covariance matrix C_X is given by the following equation :

$$C_X = J_{X,U} C_u J_{X,U}^T \quad (4)$$

where

$$J_{X,U} = \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial d} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial d} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial d} \end{bmatrix} = \begin{bmatrix} \frac{b}{d} & 0 & \frac{-ub}{d^2} \\ 0 & \frac{b}{d} & \frac{-vb}{d^2} \\ 0 & 0 & \frac{-fb}{d^2} \end{bmatrix} \quad (5)$$

We obtain

$$C_X = \begin{bmatrix} A + Bu^2 & uvB & ufB \\ uvB & A + Bv^2 & vfB \\ ufB & vfB & A + Bf^2 \end{bmatrix} \quad (6)$$

Where $A = (\frac{bp}{d})^2$; $B = m\left(\frac{b}{d^2}\right)^2$.

With the above error model, given the pointing and matching error for a stereo camera system, we can determine the covariance matrix for each 3D point in the disparity image. This covariance matrix defines the confidence we have in the accuracy of the 3D point position estimation.

The distance between a 3D point i=(x, y, z) and a given point j=(x_j, y_j, z_j) in the workspace W is given by :

$$\rho = \sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2} \quad (7)$$

To obtain the error $\triangle\rho$ in the distance between a 3D point i=(x, y, z) and a fixed point j=(x, y, z), we propagate the error from the space (x, y, z) to the ρ coordinate.

$$\triangle\rho = J_{\rho,X} C_X J_{\rho,X}^T \quad (8)$$

with

$$J_{\rho,X} = \begin{bmatrix} \frac{\delta\rho}{\delta x} \\ \frac{\delta\rho}{\delta y} \\ \frac{\delta\rho}{\delta z} \end{bmatrix}^T = \frac{1}{\rho} \begin{bmatrix} x \\ y \\ z \end{bmatrix}^T \quad (9)$$

$$\triangle\rho = \frac{Cx^2 + Dy^2 + Ez^2 + 2Fxy + 2Gxz + 2Hyz}{\delta^2} \quad (10)$$

Where :

$C = (\frac{bp}{d})^2 + m\left(\frac{ub}{d^2}\right)^2$

$D = (\frac{bp}{d})^2 + m\left(\frac{vb}{d^2}\right)^2$

$E = (\frac{bp}{d})^2 + m\left(\frac{fb}{d^2}\right)^2$

$F = uvm\left(\frac{b}{d^2}\right)^2$

$G = ufm\left(\frac{b}{d^2}\right)^2$

$H = vfm\left(\frac{b}{d^2}\right)^2$

4.2 Three dimensional tessellation of workspace

We use a 3D occupancy grid to model the environment. The workspace is discretized to uniform cubes (voxels) which are the smallest volumes of the environment model. The size of voxels determines the desired resolution of the model. Based on the captured information with the stereo vision system, the state of each voxel may be determined using a state function.

$$W_{3D} = \bigcup_{i=1}^{n} V_i, \forall i, j \in [1, n], V_i \cap V_j = \phi \quad (11)$$

Where W_{3D} is the workspace and V_i is the voxel i in the 3D model. The state of a voxel is a variable function with values in $[0, 1]$.

4.3 Voxel occupancy state observation

We propose an observation function O to convert provided stereo data to voxel occupancy state estimate. The observation on the state of a voxel is based on the mutual positions of voxels and the determined 3D points.

We consider that the resolution of the model is $2r$, i.e. the size of a voxel is $2r \times 2r \times 2r$. The observation at a voxel is estimated based on the determined$3D$ points as follow:

$$O_j(V_i) = \begin{cases} 1 & if \quad 0 \le \rho \le r - \triangle\rho \\ 1 - \dfrac{\rho}{r} & if \quad r - \triangle\rho < \rho \le r \\ 0 & otherwise \end{cases} \quad (12)$$

$O_j(V_i)$ is the observation at the voxel centered in the pixel i based on the position of the 3D point j. ρ is the euclidean distance between the center of the voxel and the 3D point j, $\triangle\rho$ is the estimate error in the distance.

The stereo information is fused into the occupancy grid using union operator to determine the final observation of

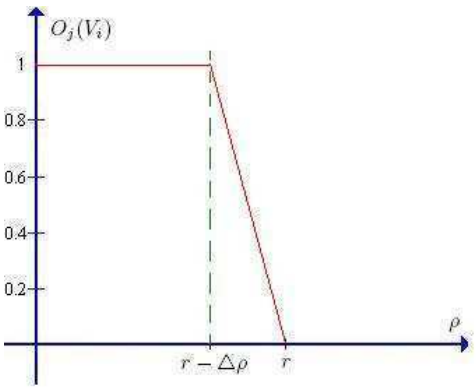

Fig. 2. Calculation of voxel state observation based on error propagation and distance to 3D points.

the voxel V_i at the instant t (the instant when the stereo images were taken).

$$O_t(V_i) = min\left(\max_j [O_j(V_i)] + \lambda \frac{N(V_i)}{6r^3}, 1\right) \quad (13)$$

$N(V_i)$ is the number of the $3D$ points inside the voxel V_i. A gratification of $\frac{N(V_i)}{6r^3}$ is given to the voxel based on the number of $3D$ points found inside it. λ is a scaling constant empirically determined. The observation value $O_t(V_i)$ reveals the degree of occupancy of the area represented by the voxel based on the input stereo information taken at the instant t.

4.4 Model update

In our approach, to update the state of each voxel, we use a *credibility* value that ranges from 0 to 1. The credibility value $k_{i,t}$ states how much we are able to trust the observation $O_t(V_i)$ of the voxel i calculated based on the stereo pair taken at the instant t. Given a time t and a new occupancy observation $O_{t+1}(V_i)$, the state of the corresponding voxel is updated following this equation:

$$\begin{cases} S_{t+1}(V_i) = (1 - k_{i,t+1})S_t(V_i) + k_{i,t+1}O_{t+1}(V_i) \\ S_{t_0}(V_i) = 0 \end{cases} \quad (14)$$

The credibility value $k_{i,t}$ is dependent on the neighborhood homogeneity of the determined voxel, the quantity of prior measurements and the age of last observation. It is unlikely to find a single occupied voxel in an otherwise empty environment, so measurements indicating homogeneous regions are more likely to be credible and we don't want to trust the very first measurements and over aged measurement at a point too much. The neighborhood homogeneity of an observation $O_t(V_i)$ is calculated using a set of voxel N in the neighborhood of V_i. In our approach, only directly neighboring voxels are regarded. The following equation gives the homogeneity of the observation at the voxel V_i at the instant t.

$$\mathcal{H}_{i,t} = \frac{\sum_{V_j \in N} |O_t(V_i) - O_t(V_j)|}{|N|} \quad (15)$$

The credibility value is given by the following equation:

$$k_{i,t} = \frac{N_{i,t}(1 - \mathcal{H}_{i,t})}{\sqrt{2\pi}} exp\left(-\frac{t - t_{last}}{2\sigma^2 t_0}\right) \quad (16)$$

Fig. 3. Office scene

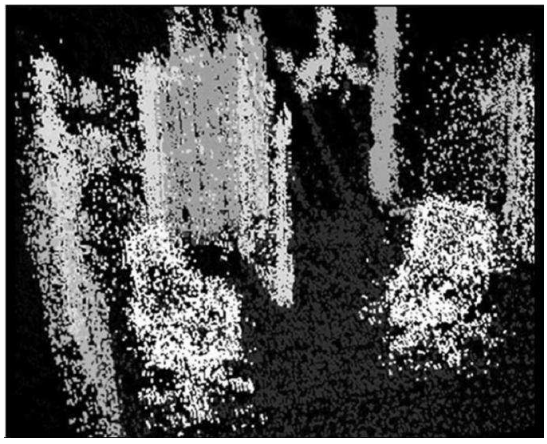

Fig. 4. Results for Voxels states representation using three stereo pairs taken from three different positions of the stereo cameras, white voxels have high occupancy state,

Where $N_{i,t}$ is the number of prior observations calculated for the voxel V_i until the instant t. t_{last} is the time of the last observation (before the instant t) calculated for the voxel V_i. σ is the age scaling constant. The age of last measurements, number of prior measurements are called *meta information* and are stored in each voxel.

Fig. 3 shows the representation of the states of voxels in a 3D scene of an office (Fig. 2). White voxels represent occupied voxels, black ones are free.

4.5 2D map generation based on projection of occupied voxels

A major drawback of occupancy grid based methods is their large memory requirement. The 3D grid map needs to be initialized so that it is at least as big as the bounding box of the mapped area, regardless of the actual distribution of map cells in the volume. In large-scale outdoor scenarios or when there is the need for fine resolutions, the memory consumption becomes prohibitive. Pruning of the 3D occupancy grid obtained by the stereo vision cameras is needed to achieve computational efficiency in real time environment, after which the 3D occupancy grid

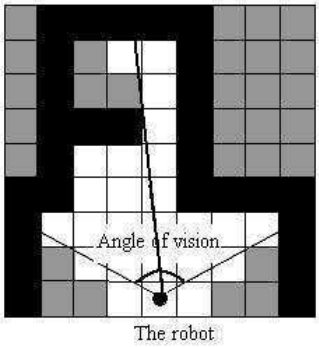

Fig. 5. 2D map building; black cells are *occupied*, white cells are *free* and gray cells are *unknown*

is scaled down to a 2D vision map, to reduce computational costs. We create a 2D map from the 3D voxels generated as described in previous sections. The 2D workspace is decomposed into cells as follow:

$$W_{2D} = \bigcup_{k=1}^{m} C_k, \forall k, l \in [1, m], C_k \cap C_l = \phi \tag{17}$$

The occupancy state of a cell is given by the maximum state of all the voxels that projects into the cell under consideration.

$$S(C_k) = \max_{V_i \in P(C_k)} (S(V_i)) \tag{18}$$

Where $S(C_k)$ is the cell state, $P(C_k)$ is the set of all the voxels that projects into the cell C_k. The occupancy state of the cells is discretized into *unknown*, *free* and *occupied* according the value of $S(C_k)$. All the cells are initially labeled as *unknown*. When a state is calculated for a cell, it is then compared to a threshold. If the state value of the cell is upper to the threshold, it is then labeled as *occupied*. All the cells between the robot and *occupied* cells are labeled *free* as shown in Figure 5.

5. EXPERIMENTAL RESULTS

To evaluate our method for environment modeling we use a benchmarking suite for occupancy grid mapping presented by Thomas Collins et al. (Collins et al. 2007). This benchmark suite encompasses an image comparison algorithm based on correlation (T. Collins et al. 2005), a direct comparison method called map scoring based on (Martin and Moravec, 1996) and a path analysis technique which tests the usefulness of a map as a means of navigation rather than treating it as if it were a picture. The correlation metric is calculated by matching the generated map with the theoretical map using the following equation:

$$C = \frac{\langle M.T \rangle - \langle M \rangle \langle T \rangle}{\sigma(M) \times \sigma(T)} \tag{19}$$

Where M is the map to be matched, T is the theoretical map, $\langle\rangle$ is the average operator and σ is the standard deviation over the area being matched. C is a percentage value that specifies the similarity of the two maps. The *map score* uses a normalization map to calculate the sum of the squared differences between corresponding cells. The map score is given by the following equation.

Fig. 6. The mobile robot Pioneer 3

Fig. 7. The test environment

$$MapScore = \sum_{m_{xy} \in M, n_{xy} \in N} (m_{xy} - n_{xy})^2 \tag{20}$$

Where m_{xy} is the value of the cell at position (x, y) in M and n_{xy} is the value of the cell at position (x, y) in the normalized map N. Map score gives a positive value representing the difference between two maps (generally the ideal map of the environment and the generated map that we are evaluating), so the lower the number, the more alike the two maps are.

The benchmarking method also calculates a *false positive* score that expresses the degree to which the paths created in the generated map would cause the robot to collide with a structural obstacle in the real world and a *false negative* score that expresses the degree to which the robot should be able to plan a path from one position to another using the generated map, but cannot because such paths are invalid in the ideal map. The method for calculating the false positives and false negatives is detailed in (Collins et al. 2007).

Experimentation consisted of testing the mapping paradigms with identical data obtained from a number of runs in a real indoor environment using Pioneer 3 (Figure 6) equipped with stereo vision cameras. Stereo vision processing and map updates are done at 5Hz and Pioneer 3 moves at 0.2 m/s. The test environment is an office in the LIRMM Laboratory (Figure 7).

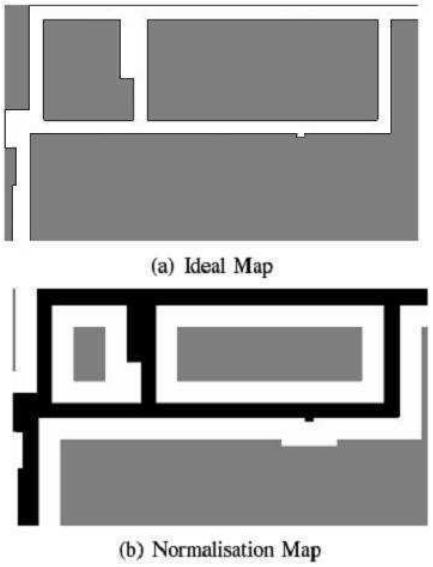

Fig. 8. Normalization of ideal map for benchmarking

Figures 8(a) and 8(b) show the ideal map and the normalized map for the test environment used for benchmarking. Figure 9 presents some illustrative maps generated by the various mapping paradigms over a single run in the test environments. The map cell size used is $25cm$.

Table I presents the results for the comparison of our method with five mapping paradigms. We calculate the correlation metric, map score metric for all cells and for the occupied cells, the false positives and negatives for the different algorithms.

Table 1. Comparison of our results with the results of the different mapping algorithms using the Collins et al. benchmarking method

	Corr.	A.M.Sc	O.M.Sc	F.Pos	F.Neg
key	A	B	C	D	E
Our method	53.56	17.21	19.63	44.47	14.93
Rank	1	1	2	1	1
Mor. & Elf. 85	39.18	33.73	23.56	72.84	21.34
Mat. & Elf. 88	40.69	28.27	24.82	69.17	25.91
Thrun 93	38.34	25.97	29.71	77.13	27.93
Konolige 97	40.54	20.25	19.69	63.45	22.64
Thrun 01	50.13	18.56	17.39	50.15	16.68

5.1 Analysis of the correlation results

As shown in Table 1 (key A) our method achieved the highest correlation followed by the Thrun's forward model based paradigm. Thrun's 1993 paradigm has the lowest correlation of the systems tested. According to Collins et al. (Collins et al. 2007) this can be attributed to two causes. First, paradigms that have low correlation cost have a tendency to overestimate free space as can be clearly seen from Figure 9 and also have a tendency to model the extremities of the sensors as being occupied. In our method, inputs are issued from stereo cameras, the problem of sensor extremities does not exist. As shown in Figure 9, the sensors extremities are modeled as unoccupied. The problem of the overestimation of the free space is due to the updating mechanisms. In fact, according to (Collins et al. 2007), the Bayesian

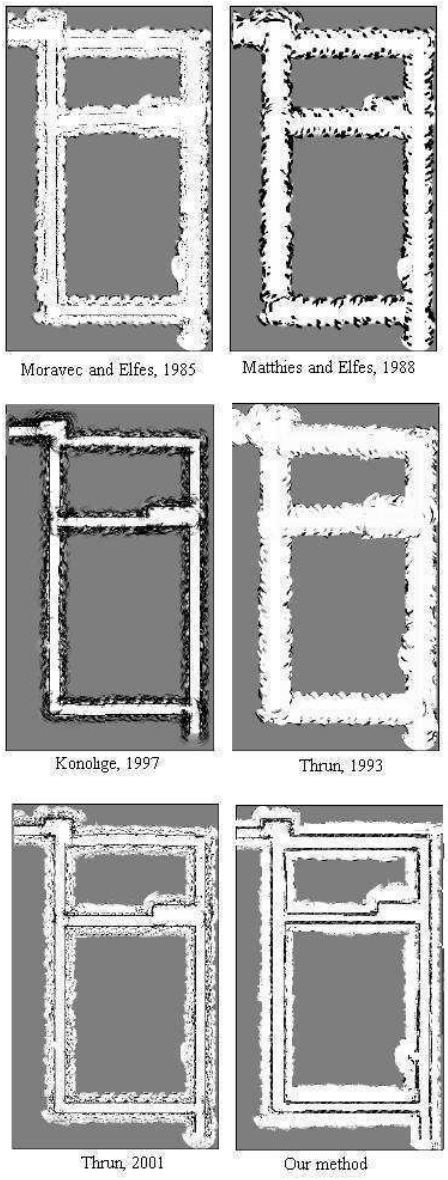

Fig. 9. Illustrative occupancy grid maps generated during experimentation of the different mapping algorithm. Environmental size: $(15 \times 20\ m^2)$.

update makes the paradigm susceptible to fluctuations in occupancy values when used in conjunction with an approach that has a penchant for over estimation of occupied/empty space. In our approach, we have used a credibility based update method. To update the model, we take into account, the age of the last observation and the homogeneity of the current observation in a local neighborhood. As shown in Figure 9, our update paradigm don't suffer from free space over estimation problem.

5.2 Analysis of the map score results

A lower map score indicates a less of a difference between the generated map and the theoretical map. Table I (B) presents the results for the map score all metric. As can be seen, our method has the best performance achieving the lowest map score followed by Thrun's 2001 method and Konolige's 1997 method. The map score

metric compares the generated maps with a theoretical map. The reasons for this performance are the same outlined for the correlation results. Table I (C) presents the results for the map score occupied cells metric. Thrun's 2001 method achieved the best performance followed by our method. This is because we use a relatively high threshold to label a cell as occupied. In fact, the first criteria to be satisfied in our approach is the robustness of the path. For this purpose, all the cells are initially labeled as unknown. The status of a cell is modified to free or occupied only when we have sufficient data that favor this change.

5.3 Analysis of the "false positives" results

Mapping approaches that have poor performance regarding false positives metric are those that have a tendency to update free-space too strongly (Collins et al. 2007). Our approach achieved the first performance as shown in Table I (key D) followed by Thrun's 2001 paradigm. This is because the structure of the occupied space in maps generated by the 1993, 1988 and 1985 paradigms cause a number of inconsistent paths to be created due to the tendency to render areas that reach past the occupied cells to the extremity of the sensor beam as being unoccupied. This subsequently causes the creation of possible paths on either side of the correctly identified environmental obstacles, which are not possible in the actual environment.This indicates that paths generated by our mapping system are robust and usable.

5.4 Analysis of the "false negatives" results

The false positives benchmark was concerned with determining the usability of the map as a basis for safe robot navigation, this metric presents the percentage of false negative paths in the map and is concerned with the usability of the map as a basis for planning a path in the real world environment. The false positives metric informs about the number of paths that cannot be completed in the generated map but can be completed in the theoretical map. Table I (key E) shows that our approach achieved the best performance giving the best map for path planning.

6. CONCLUSION

In this paper, a 2D map building algorithm is proposed based on binocular stereo vision. The algorithm is highly robust and meets the need of navigation in real environment. An intrinsically uncertain 3D representation of the environment based on error propagation is used. A 3D occupancy grid is used to model the environment. A new update method based on a proposed credibility value is used to update environment model. Finally, the 3D occupancy grid is scaled down to a 2D navigation map. Experimental results have been reported to illustrate the satisfactory performance of the proposed method. The obtained maps were quite accurate in real environment and permit to identify obstacles and free space.

REFERENCES

Albouy, B., Koenig, E., Treuillet, S., & Lucas, Y., (2006) "Accurate 3D structure measurements from two uncalibrated viewspositions", *Advanced Concepts for Intelligent Vision Systems* 4179, 1111–1121.

Badino, H., Franke, U., & Mester, R. (2007). "Free space computation using stochastic occupancy grids and dynamic programming". *Workshop on Dynamical Vision, ICCV*, Rio de Janeiro, Brazil.

Balasuramanian, R., Das, S., and Swaminathan, K., (2000). "Error analysis in reconstruction of a line in 3D from two arbitrary perspective views", *International Journal of computer vision and mathematics* 78, 191–212.

Blostein, D. S. & Huang, S. T., (1987). " Error analysis in stereo determination of 3D point positions", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 9(6), 752–765.

Boyer, K. L. Wuescher, D. M. & Sarkar, S. (1991). "Dynamic Edge Warping: An Experimental System for Recovering Disparity Maps in Weakly Constrained Systems", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21,No. 1 January/February 1991.

Braillon, C. Usher, K. Pradalier, C. Crowley, J. L. Laugier, C. (2006). "Fusion of stereo and optical flow data using occupancy grids", *Proc. of the IEEE Int. Conf. on Intelligent Transportation Systems*, Toronto, CA, 2006.

Collins, T. Collins, J. J. Ryan, C. (2007). "Occupancy Grid Mapping: An Empirical Evaluation", *Proceedings of the 15th Mediterranean Conference on Control and Automation,* July 27-29, Athens, Greece, 2007.

Collins, T. Collins, J. J. O'Sullivan, S. and Mansfield, M. (2005) "Evaluating techniques for resolving redundant information and specularity in occupancy grids", *Advances in Artificial Intelligence*, pp. 235–244, 2005.

Elfes, A. E. (1987). "Sonar-based Real-World Mapping and Navigation", *IEEE Journal of Robotics and Automation*, Vol. RA-3, No 3, June 1987, pp. 249-265, 1987.

Elfes, A. E. (1989a). "Occupancy Grids: A Probabilistic Framework for Robot Perception and Navigation". *PhD thesis*, Department of Electrical and Computer Engineering, Carnegie Mellon University, 1989.

Elfes, A. E. (1989b). "Using Occupancy Grids for Mobile Robot Perception and Navigation", *Computer Magazine*, Vol. 22, No. 6, June 1989, pp. 46–57, 1989.

Elfes, A., (1992), "Multi-source spatial data fusion using Bayesian reasoning". *In: Abidi, M.A., Gonzalez, R.A. (eds.) Data fusion in robotics and machine intelligence, ch. 3.* Academic Press, New York .

Faugeras, O. D. (1993) "Three-Dimensional Computer Vision: A Geometric Viewpoint". *MIT Press*, 1993.

Franco, J. S. Boyer, E. (2005). "Fusion of Multi-View Silhouette Cues Using a Space Occupancy Grid", *Technical report, no 5551*, INRIA, April 2005, 20 pages.

Gambino, F., Oriolo, G., Ulivi, G., (1996). "Comparison of three uncertainty calculus techniques for ultrasonic map building". *In: Proc. SPIE Int. Symp. On Aerospace/Defense Sensing and Control*, vol. 2761, pp. 249–260.

Ghazouani, H. Tagina, M., Zapata, R. (2010). "A Theory of Possibility for Reliable Correspondence Search" *International Journal of Signal and Image Processing*, Vol.1–2010/Iss.4, pp. 232–237.

Guadarrama, S., Ruiz-Mayor, A., (2010). "Approximate robotic mapping from sonar data by modeling perceptions with antonyms", *Information Sciences: an Inter-*

national Journal , Volume 180 Issue 21, November, 2010.

Hirschmüller, H. Innocent, P. R. and Garibaldi, J. M. (2002). "Fast, unconstrained camera motion estimation from stereo without tracking and robust statistics," i*n Seventh International Conference on Control, Automation, Robotics and Vision*, (Singapore), pp. 1099–1104, 2–5, December 2002.

Howard, A. & L. Kitchen (1997). "Sonar mapping for mobile robots". *Technical Report 96/34*, Department of Computer Science, University of Melbourne, March 1997.

Jianxi, Y., Jianting, L., and Zhendong, S., (2008). "Calibrating method and systematic error analysis on binocular 3D position system", *Proceedings of the 6th International Conference on Automation and Logistics*, China, 2310–2314.

Kamberova, G. and Bajcsy, R., (1998). "Sensor errors and the uncertainties in stereo reconstruction", *Proc. IEEE Workshop on Empirical Evaluation Techniques in Computer Vision in conjunction with CVPR 98*, 96–116

Kohara, K., Suganuma, N., Negishi, T., & Nanri, T.(2010) "Obstacle Detection Based on Occupancy Grid Maps Using Stereovision System " *International Journal of Intelligent Transportation Systems Research*, (2010) 8:85–95.

Konolige, K. (1997). "Improved occupancy grids for map building," *Autonomous Robots*, no. 4, pp. 351–367, 1997.

Koren, Y. & J. Borenstein (1991). Potential eld methods and their inherent limitations for mobile robot navigation. *In IEEE International Conference on Robotics and Automation (ICRA'91)*, pp. 1398–1404.

Lanthier, M., D.Nussbaum, and A.Sheng (2004, August). "Improving Vision-Based Maps by using sonar and infrared data". *Robotics and Applications, IASTED 2004*.

Lategahn, H. Derendarz, W. Graf, T. Kitt, B. Effertz, J. (2010). "Occupancy grid computation from dense stereo and sparse structure and motion points for automotive applications ", *2010 IEEE Intelligent Vehicles Symposium (IV)*, 1931-0587, San Diego, CA 21–24 June 2010.

Martin, M. and H. Moravec (1996). "Robot Evidence Grids", *Technical report CMU–RI–TR–96–06*, The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, March 1996.

Matthies, L. and Elfes, A. (1988). "Integration of sonar and stereo range data using a grid-based representation," *in Proceedings of the 1988 IEEE International Conference on Robotics and Automation*, 1988.

Moravec H. and Elfes, A. (1985). "High resolution maps from wide angle sonar," *in Proceedings of the 1985 IEEE International Conference on Robotics and Automation*, 1985.

Moravec, H., (2001), "DARPA MARS program research progress". *Carnegie Mellon University* http://www.frc.ri.cmu.edu/hpm/talks/Report.0107.html

Oniga, F., Nedevschi, S., Danescu, R., Meinecke, M., (2009) "Global map building based on occupancy grids detected from dense stereo in urban environments". *IEEE 5th International Conference on Intelligent Computer Communication and Processing, 2009. ICCP 2009.* 111-117 Cluj-Napoca, 27–29 Aug. 2009.

Oriolo, G., Ulivi, G., Vendittelli, M., (1999). "Real-time map building and navigation for autonomous robots in unknown environments". *IEEE Transactions on Systems, Man, and Cybernetics 5.*

Park, S. and Subbarao, M., (2005). "A multiview 3D modeling system based on stereo vision techniques", *Machine Vision and Applications*, 16(3), 148–156.

Ramakrishna, R. S. and Vaidvanathan, B., (1997). "Error analysis in stereo vision", *Computer Vision ACCV*, 98 1351, 296–304.

Ribo, M., Pinz, A., (2001). "A comparison of three uncertainty calculi for building sonar-based occupancy grids". *International Journal of Robotics and Autonomous Systems* 35, 201–209.

Rivera-Rios, A. H., Shih, F., and Marefat, M., (2005). "Stereo camera pose determination with error reduction and tolerance satisfaction for dimensional measurements", *Proceedings of the 2005 IEEE Int. Conf. on Robotics and Automation*, Barcelona, Spain , 423–428.

Schmid, M.R. Maehlisch, M. Dickmann, J. Wuensche, H. J. (2010). "Dynamic level of detail 3D occupancy grids for automotive use ", *2010 IEEE Intelligent Vehicles Symposium (IV)*, 269–274, San Diego, CA, 21–24 June 2010.

Tirumalai, A. P. Schunk, B. G. and Jain, R. C. (1995). "Evidential reasoning for building environment maps," *IEEE Transactions on System, Man and Cybernetics*, vol. 25, pp. 10–20, January, 1995.

Thrun, S. (1993). "Exploration and model building in mobile robot domains." *in Proceedings of IEEE International Conference on Neural Networks*,. Seattle, Washington, USA: IEEE neural Network Council, 1993, pp. 175–180.

Thrun, S. (2001). "Learning occupancy grids with forward models," *in Proceedings of the Conference on Intelligent Robots and Systems (IROS'2001)*, 2001.

Yang, Z. and Wang, Y. F.(1996). "Error analysis of 3D shape construction from structured lighting," *Pattern Recognition*, 29, 189–206.

Yoon K. and Kweon I. (2006). Adaptive support-weight approach for correspondence search. IEEE Trans. *PAMI*, 28(4):650–656, 2006.

Youngblood, G. M.; Holder, L. B. & Cook, D. J. (2000). "A Framework for Autonomous Mobile Robot Exploration and Map Learning Through the Use of Place-Centric Occupancy", *ICML Workshop on Machine Learning of Spatial Knowledge*, 2000.

Zhang, Z. Hou, C. Shen, Yang, J. (2009) "An Objective Evaluation for Disparity Map based on the Disparity Gradient and Disparity Acceleration", *International Conference on Information Technology and Computer Science*, 2009.

AUTHORS PROFILE

Haythem Ghazouani received his Master's Degree in Computer Science from the National School for Computer Studies of Tunis, Tunisia, in 2006. His research interest includes soft computing, robotics, computer vision and artificial intelligence. He is currently a Ph.D. student in the University of Montpellier II, France and the the National School for Computer Studies of Tunis, Tunisia. His current research interest includes mobile robot navigation using stereo vision, robot cooperation, map building and occupancy grid.

Moncef Tagina is professor of Computer Science at the the National School for Computer Studies of Tunis,

Tunisia. He received the Ph.D. in Industrial Computer Science from Central School of Lille, France, in 1995. He heads research activities at LI3 Laboratory in Tunisia (Laboratoire d'Ingénierie Informatique Intelligente) on Metaheuristics, Diagnostic, production, Scheduling and robotics.

René Zapata is professor at the the University of Montpellier II, France. He heads research activities at LIRMM Laboratory (Laboratoire d'Informatique, de Robotique et de Micro Electronique de Montpellier) on humanoïd robots, robot planification, vision, map building and localization.