Simultaneous Localization and Map building paradigm based on omnidirectional stereoscopic vision

Cyril DROCOURT, Laurent DELAHOCHE, Eric BRASSART, Cyril CAUCHOIS
CREA – IUT Informatique, Avenue des facultés, 80000 Amiens – FRANCE
Cyril.Drocourt@u-picardie.fr , Laurent.Delahoche@u-picardie.fr

Abstract

This paper deals with the localization and map building paradigm in an unknown indoor environment. We propose an exploration method based on the use of the sensorial data provided by an omnidirectional stereoscopic vision system. The first part of our study is linked with the problem of sensorial model construction with two omnidirectional images obtained by a rigid translation along a rail of our SYCLOP sensor. We propose an approach based on the fusion of several criteria which is realized according to Dempster-Shafer rules. The second part is devoted to the matching problem of the stereo sensorial model with an environment map integrating all the previous primitive observations. We propose two matching approaches based on different selection criterion: the Hausdorff distance and the cumulated cartesian distance. The third part presents our incremental map building paradigm based on the hypothesis of a non a priori knowledge. We deal with the problem wish consists in allowing a robot to localize itself and to construct concurrently a representation of its environment. In this part we discuss the problem of interaction between the localization stage and the mapping stage.

I.Introduction

The map building problem is preponderant for the increase of mobile robot autonomy [1] [2] [3]. It consists in managing a coherent representation of the environment along a robot's displacement. The mapping stage is directly correlated to the localization stage: the coherence and the robustness of the map updating is linked to the robustness of the position estimation. For a dynamic map building it is necessary to localize the robot in relation with already known map elements. There are different types of maps mainly based on the nature of sensorial data and the representation requirements of localization. We can distinguished mainly two kinds of map representation: the metric approaches and the topological approaches. The first approach consists in managing the notion of distance. We can find mainly two types of mapping paradigm to take into account the notion of distance. The first paradigm consists in computing a cartesian representation of the environment which generally used the Extended Kalman filtering. This is mainly what Crowley has pointed out by using the Kalman Filtering technique to build the environment global map and to localize the robot [6]. The fusion of dead reckoning and ultrasonic data is thus realized. Kalman Filtering is also used by Leonard and Durrant-Whyte [5] to realize the dynamic building of the environment map with ultrasonic data. More recently, and linked to the previous works, the SLAM algorithm (simultaneous localization and map building) [2] permits to build a cartesian map of an outdoor environment with dead-reckoning sensors, laser range and bearing information. In the same way, and based on the SLAM algorithm, an extended Kalman filter (EKF) maintains an estimate of map features in addition to vehicle state, with the use of a millimeter wave radar [4]. An omnidirectional sensor can be use in connection with this kind of approach [3] [1]. The use of the Extended Kalman Filtering in the localization and mapping process set the problem of divergence linked to the dead-reckoning prediction used for the linearization.

The second approach to manage metric maps is occupancy grid maps, which were originally proposed in [7] [8] and which have been employed successfully in numerous mobile robot systems. Occupancy grids are designed to estimate the occupancy of all cells in the environment. This type of representation is also used by Boreinstein in [9]. More recently Dieter Fox introduces in [10] a general probabilistic approach to concurrent mapping and localization. This method poses the mapping problem as a statistical maximum likelihood problem, and devises an efficient algorithm for search in likelihood space. In the same way that previous approaches, the proposed paradigm in [10] addresses the problem of using occupancy grid maps for path planning in highly dynamic environments. This grid approach permits to solve the problem previously hightlighted to the EKF mapping methods. However, a major drawback of occupancy grids is caused by their pure sub-symbolic nature: they provide no framework for representing symbolic entities of interest such as doors, desks, etc [10].

The second category of map representation is the topological one. This approach consists in determining and managing the location of significant places in the environment along with an order in which these places were visited by the robot. In the topological mapping step, the robot can generally observes whether or not it is at a significant place. The definition of significant places can be linked for example to the notion of "distinctive places" in the Spatial Semantic Hierarchy proposed in [11], and the notion of "meetpoints" in the use of Generalized Voronoi Graphs proposed in [12]. This kind of method is interesting to use in complement with an occupancy grid, in order to take into account the semantic aspect.

We propose a Simultaneous Localization and Map building paradigm based on a geometric approach. This paradigm takes into account the hypothesis of a non a priori knowledge. Our method is not based on the use of an Extended Kalman Filter but on a matching stage based on different selection criterion: the Hausdorff distance and the cumulated cartesian distance. In the first part we deal with the problem of the stereoscopic sensorial model construction. The second part is devoted to the localization stage. In the third part we propose a localization and map building paradigm based on multi-criteria approach for the matching stage and on the use of the recursive least square method for the primitive coordinates computation. Finally, we present our reconstruction experimental results obtained on a large indoor environment.

II. Sensorial primitive

The stereoscopic omnidirectional sensor put on our mobile robot SARAH is based on the rigid translation of an omnidirectional vision system SYCLOP used in our laboratory [3]. The SYCLOP system is composed of a conic mirror and a CCD camera (Figure 1). The rigid translation is been made thanks to two horizontal rails which allow a precise straight move in the horizontal SYCLOP sensor plan.

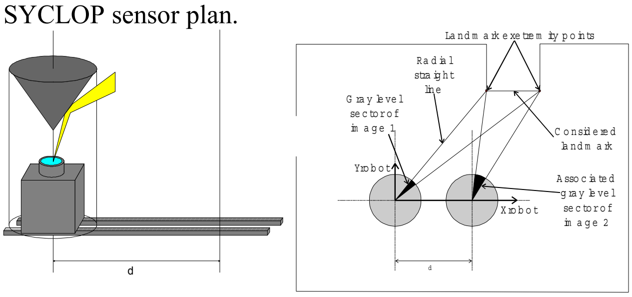

Figure 1 : Principle of the stereoscopic omnidirectional sensor

Thus the system insures the acquisition of two omnidirectional images of the environment within 40 centimeters of one another. To calculate the coordinates (x,y) of a point (corresponding to a vertical landmark like edges, corners …) it's necessary to know the « landmark's signature » in the two omnidirectional images. This signature is characterized by a radial straight line on each image: after a matching stage, these two radial straight lines allow to compute by triangulation the point coordinates in the robot's reference.

We take into account the characteristics of the considered lines primitives: they are radial and converge to the center of the cone. Then we have decided to work on pixels which are included in several concentric circles centered on the top of the cone. We work finally on one grey level circle computed with the mean of grey level pixels belongs of several circles. The mean value is computed in connection with radial directions. To extract the radial straight lines, we apply a one dimension Sobel gradient filter on the circle computed with this mean [13].

The stereo sensorial model is obtained after a matching stage. This stage consists in associating a grey level sector (characterized from two radial straight lines) of the first omnidirectional image with its corresponding sector on the second image (Figure 3). Since the sectors don't following one another in the same order on the two images, we have selected 4 comparison criteria for each sector (Figure 2):

- inclination of the approximate straight lines corresponding to the set of sector gray level,
- average of the gray level determined by the set of sector grey level,
- standard deviation of the gray level determined by the set of sector grey level,
- the geometrical criterion linked to the sector, which can described like a "pseudo epipolar criterion", we can see than in Figure 4 where α is necessary inferior to β [13].

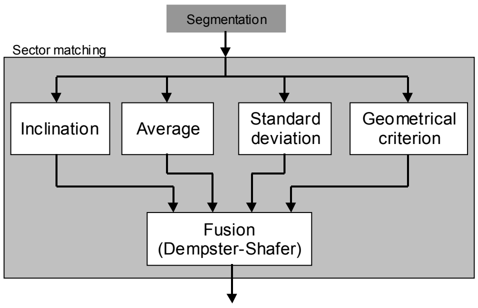

Figure 2 : Global matching sector algorithm

The merging of these "heterogeneous" criteria is based on the use of the Dempster-Shafer's rules [13][15].

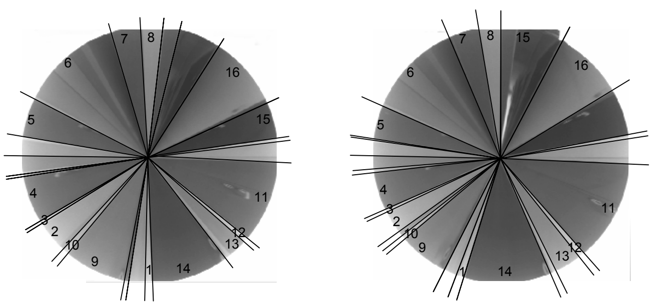

Figure 3 : Segmentation and final sector matching for the two images corresponding to a stereoscopic acquisition

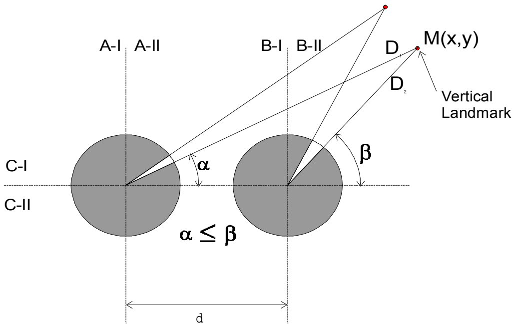

Figure 4 : Geometric constraints for the sector matching

The final stage which allows us to obtain the sensorial model consists in computing the points coordinates by triangulation (1).

$$x = \frac{d \times \tan(\beta)}{\tan(\beta) - \tan(\alpha)} \qquad y = \frac{d \times \tan(\beta) \times \tan(\alpha)}{\tan(\beta) - \tan(\alpha)} \qquad (1)$$

The stereo sensorial model obtained is composed of point primitives corresponding to a vertical landmark of the robot's environment (Figure 5).

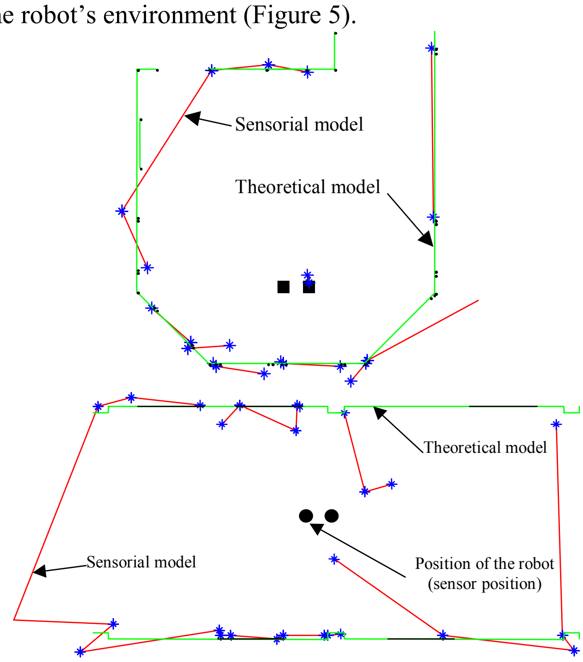

Figure 5 : Stereo sensorial model in two different environments (hall and corridor) compared with the real environment (theoretical model).

We show on figure 5 a superposition of the sensorial model with a theoretical representation of the environment (with an a priori known map). The figure 5 allows to highlight the robustness of the stereoscopic sensorial model.

III. localization approaches

The next stage is the localization one: we can only update the environment map if the stereo sensorial model is matched with the previous merged stereo models. The localization stage can be declined like a matching of two set of points: the set of points of the stereo sensorial model and a set of points belonging to the environment map at a time t. The map will be constructed with the sensorial models obtained until time t-1. This problem is classical notably in stereo vision: it consists in finding the best matched configuration linked to a selection criteria like for example the Hausdorff distance [16].

The Hausdorff distance between two sets, $A=\{a_1,\dots,a_n\}$ and $B=\{b_1,\dots,b_n\}$, where a_i, b_j are points, is given by:

$$H(A,B) = \max(h(A,B), h(B,A))$$

Where

$$h(A,B) = \max_{a \in A} \min_{b \in B} \|a - b\|$$

and $\| \ \|$ is the Euclidian norm.

The function h(A,B) is called the direct Hausdorff distance from A to B. If h(A,B)=d, then each point of A must be within distance d of some points of B, and there also are some points of A that is exactly distance d from the nearest point of B.

The Hausdorff distance H(A,B) is the maximum of h(A,B) and h(B,A), thus it measures the degree of mismatch between two sets, by measuring the distance of the point of A that is the farthest from any point of B and vice versa.

In order to estimate a position which minimizes the Hausdorff distance, it's necessary to calculate this distance for all possible positions of the robot (figure 6). To reduce the complexity of this stage, a reduction of computation is necessary. To do it, we use the dead-reckoning pose estimation. We obtain the pose and the error domain with the classical equations linked to the use of odometers [13]. In our matching problem, we determine a domain of possible absolute pose with an overcharge of the ellipse domain error which we assimilate as a circle. The radius of this circle is calculated with the length of the major axis of the ellipse error quantification.

To compute the Hausdorff distance we manage a grid, representing the different possible positions of the robot (Figure 6). This grid includes the previous circle. For each (x,y) cells and for different possible values of angle (inside the interval given by dead-reckoning), we calculate the Hausdorff distance between points of the sensorial model and points of the environment map. Finally, the minimal Hausdorff distance gives the pose of the robot.

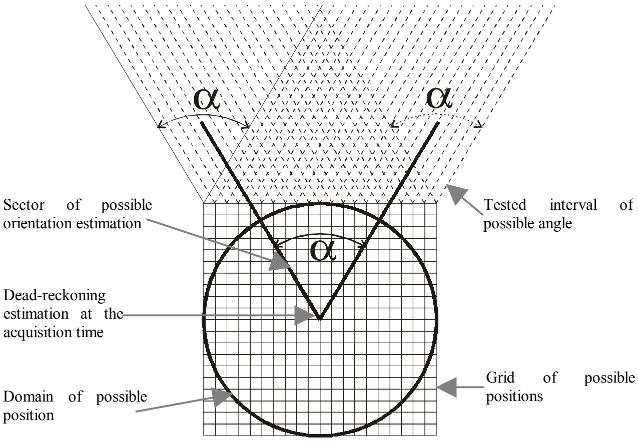

Figure 6 : Grid of possible position for the robot

We have tested this pose estimation approach on several stereo acquisitions. We can note this method is not adapted for two reasons:

- There are too few of points in the sensorial model (and also in the map)
- The Hausdorff distance is very sensible to noise

The Hausdorff distance gives good results with lot of points not blemished of noise: this does not correspond systematically to our work hypothesis.

The most robust convergence criterion that we have kept for the matching algorithm is based on the Cartesian distance between each sensorial model point and the nearest one of the environment map.

The principle of our localization method is finally as follows :

- we consider 2 points in the sensorial model
- we look for 2 points of the theoretical model corresponding to the 2 chosen points in connection with the Cartesian distance.
- Lastly we have to calculate, for each remaining sensorial model point, the minimal distance which separates it from a point of the theoretical model (incremental constructed map). The cumulated distance (sensorial point-theoretical point) allows to choose the best position. The selection criterion permits to choose the final solution is then the minimum cumulated distance error.

The first stage of this algorithm is strongly combinatorial. The first amelioration to reduce this complexity consists in using only the two points from the matching of an identical sector, which permits to reduce the number of possible couples to *n*.

Moreover in order to reduce the complexity of this stage, we use also the position estimation and its associated error domain given by dead-reckoning. Then the amelioration of the initial matching algorithm consists in ejecting the combinations of sensorial points couples and theoretical points which generate a position out of the domain of possible positions. Finally, we obtain a good algorithm of localization (Figure 7).

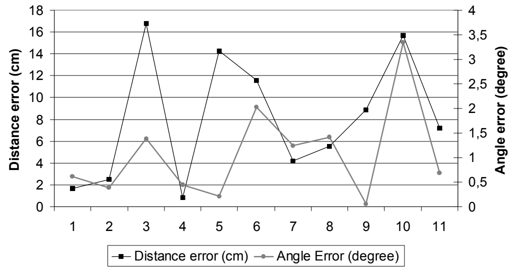

Figure 7 : Orientation and Cartesian error obtained on 11 absolute configuration estimations of the robot

IV. Map building approach

The next stage consists in updating the environment map with new observations. This stage is linked to the previous one: the fusion of the sensorial model with the map can be made only if the matching stage is achieve. The updating stage consists in the management of two cases: (1) merge a sensorial primitive with a map one (2) initialize a new primitive in the map.

Then it is necessary to merge information of the sensorial model acquired at the step n with the map built with n-1 previous acquisitions.

IV.1 - The fusion stage

The difficulty of this stage is linked to an important number of points which can be erroneous. The position of a sensorial model point is obtained from the information of the two angles (Figure 1): if only one observation is corrupted with an error, the position of the point is false even if the second observation is valid. It is why we decided to merge the two angles and not directly the information of point.

In this optic, we associate the angular errors committed by the two sensors with two functions. We get thus with matching functions the two Basic Probability Assignment (B.P.A.) associated to the fusion of an observation with an existing point of the map (Figure 8).

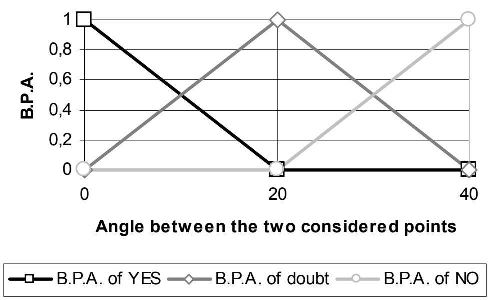

Figure 8 : Matching functions for the point fusion stage

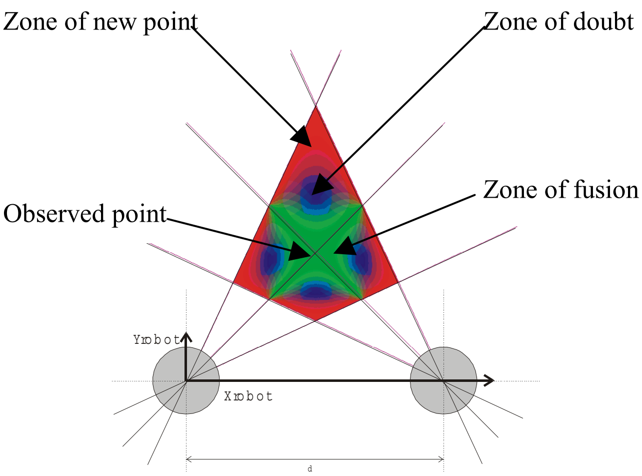

Figure 9 : BPA of the fusion area

These two B.P.A. are merged with Dempster-Shafer [15] rules to decided if the fusion is possible or not (Figure 9).

At this level, the problem is to math the s points of the sensorial model with p points of the theoretical model (environment map). To do it we use the Dempster–Shafer theory: for each point of the sensorial model S_i, we apply the following algorithm:

- The frame of discernment Θ is composed by the p points of the theoretical model and also by an element noted * which means that the point S_j cannot be matched (it's a new point). So : $\Theta=\{P_1,P_2,\ldots,P_p,*\}$
- The matching criterion is the fusion of the two differences of angle.
- For each point ***p*** of the map, we compute:
  - $m_i(P_i)$ the mass associated with the proposition "The point P_i is matched with the point S_j"
  - $m_i(\overline{P_i})$ the mass associated with the proposition "The point P_i is not matched with the point S_j"
  - $m_i(\Theta_i)$ the mass represented the ignorance concerning the point P_i

After the treatment of all points Pi, we have ***p*** triplets :

$m_1(P_1)$ $m_1(\overline{P_1})$ $m_1(\Theta_1)$

$m_2(P_2)$ $m_2(\overline{P_2})$ $m_2(\Theta_2)$

… … …

$m_p(P_p)$ $m_p(\overline{P_p})$ $m_p(\Theta_p)$

We compute the Dempster rule of combination on these triplets and we get $m^j(P_1)$, $m^j(P_2)$, …, $m^j(P_p)$, $m^j(*)$ and $m^j(\Theta)$ where :

- $m^j(P_i)$ is the mass on the proposition "The point S_j is matched with the point P_i"
- $m^j(*)$ is the mass on the proposition "The point S_j is a new point"
- $m^j(\Theta)$ is the mass on the proposition "We don't know anything about the matching of the point S_j"

In this matching case, Gruyer [14] shows that we can obtain these condensed formulas:

$$K=\frac{1}{\prod_{i=1,p}\left(1-m_i(P_i)\right)\times\left(1+\sum_{i=1}^{p}\frac{m_i(P_i)}{1-m_i(P_i)}\right)} \tag{2}$$

$$m^j(P_i)=K\times m_i(P_i)\times\prod_{\substack{k=1,p\\k\neq i}}\left(1-m_k(P_k)\right) \tag{3}$$

$$m^j(*)=K\times\prod_{k=1,p}m_k(\overline{P_k}) \tag{4}$$

$$m^j(\Theta)=K\times\left(\prod_{k=1,p}\left(m_k(\Theta_k)+m_k(\overline{P_k})\right)-\prod_{k=1,p}m_k(\overline{P_k})\right) \tag{5}$$

We compute these values for each point of the sensorial model, and we obtain the following table:

$m^1(P_1)$	…	$m^1(P_i)$	…	$m^1(P_p)$	$m^1(*)$	$m^1(\Theta)$
…		…		…	…	…
$m^j(P_1)$	…	$m^j(P_i)$	…	$m^j(P_p)$	$m^j(*)$	$m^j(\Theta)$
…		…		…	…	…
$m^s(P_1)$	…	$m^s(P_i)$	…	$m^s(P_p)$	$m^s(*)$	$m^s(\Theta)$

To correctly match the points of the sensorial model with the points of the map, we apply the following algorithm:

- We find in the previous table, the maximum value t, this value represent "the sensorial point t is matching with the point q" or "the sensorial point t is a new point".
- We suppress the line of the t point in the table, and if is not a new point, we suppress the column that contain the q point.
- We reiterate this algorithm until all the theoretical points are unmatched.

IV.٢ - The incremental map building

Our incremental map building algorithm is based on the exploitation of the angular measures. The data which allow to estimate the landmark positions to an acquisition n are, with $k=(n\times 2)-1$ and $\theta c_k=\theta c_{k+1}$:

- The two positions of observation $[xc_k \quad yc_k \quad \theta c_k]^T$ and $[xc_{k+1} \quad yc_{k+1} \quad \theta c_{k+1}]^T$
- The two azimuth angles (ϕ_k^i,ϕ_{k+1}^i) of the considered landmarks *i* in the robot's reference.

The coordinates (xb^i,yb^i) of the considered landmark i in the map reference are directly got from the following equation (Figure 10):

$$\tan(\theta c_k + \phi_k^i) = \frac{yc_k - yb^i}{xc_k - xb^i} \quad (2)$$

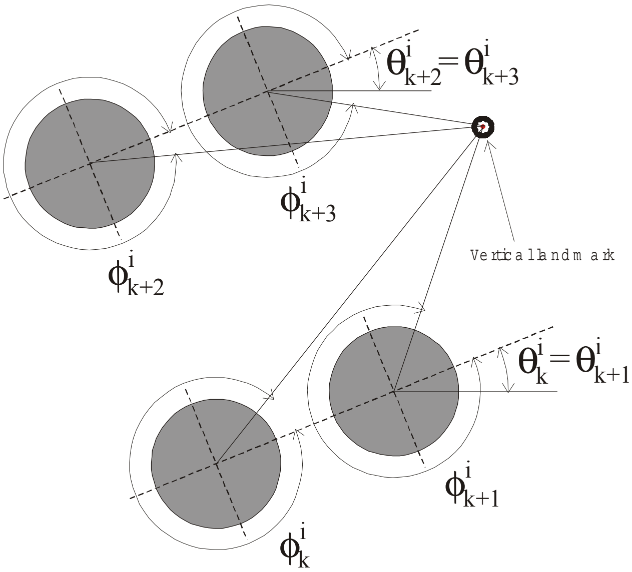

Figure 10 : Position estimation of a landmark

This equation system is overdetermined. To estimate the parameters (xb^i, yb^i) incrementally we use the recursive least squares method:

$$\begin{aligned} \underline{\hat{a}}_k &= \underline{\hat{a}}_{k-1} + K_k(y_k - \underline{x}_k^T \underline{\hat{a}}_{k-1}) \\ P_k &= P_{k-1} - K_k \underline{x}_k^T P_{k-1} \\ K_k &= P_{k-1}\underline{x}_k(1 + \underline{x}_k^T P_{k-1}\underline{x}_k)^{-1} \end{aligned} \quad (3)$$

where a_k is the state vector, y_k the observation vector and x_k represents the known parameters. To apply the recurrence equations (3) it is necessary to express the equation (2) under the following form :

$$\underline{x}_i^T \underline{a} = y_i \quad (4)$$

which gives us the equation (5) :

$$\begin{bmatrix} \tan(\alpha_k^i) & -1 \end{bmatrix} \begin{bmatrix} xb^i \\ yb^i \end{bmatrix} = xc_k \tan(\alpha_k^i) - yc_k \quad (5)$$

where $\alpha_k^i = \theta c_k + \phi_k^i$

Moreover, the recursive least squares method allows us to obtain an estimation of the error domain associated with a landmark position $\begin{bmatrix} xb^i & yb^i \end{bmatrix}^T$.

Then, the problem of noisy points is taken into account with this approach. When a point has been observed only one time and no more observed in the next acquisitions, it is suppressed automatically.

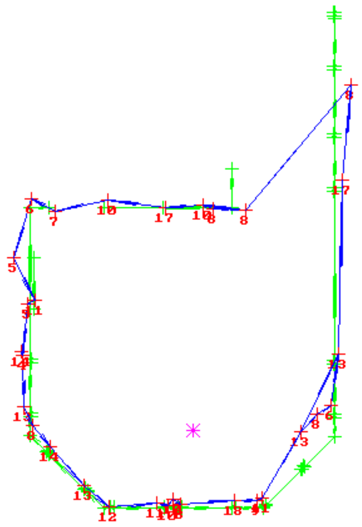

Figure 11 : Map build after 8 acquisitions

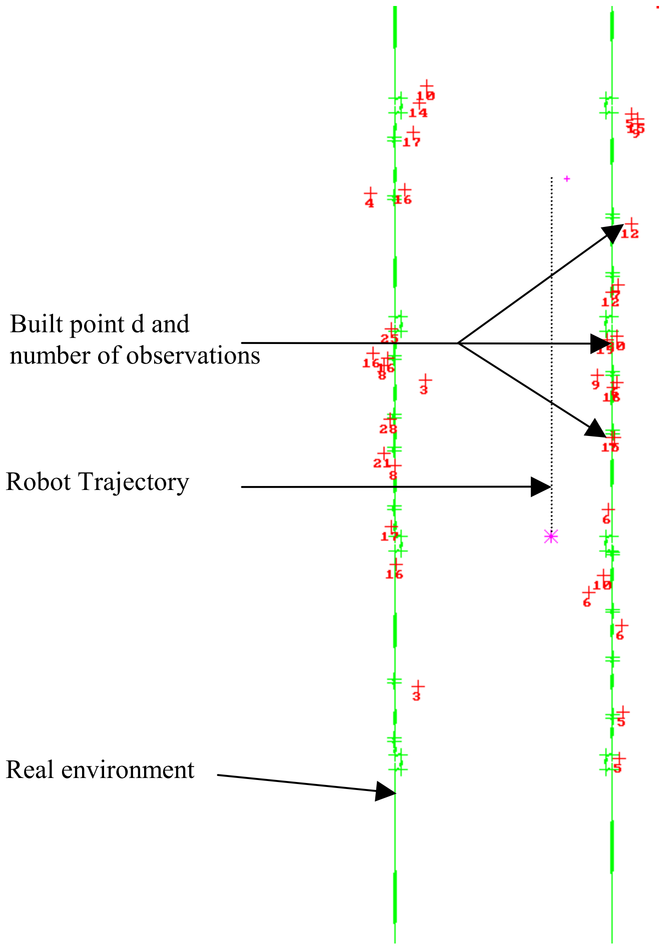

Figure 12 : Map built after 20 acquisitions

Finally, we obtain a robust theoretical environment at the n stage, this is important for localization at stage n+1. Figure 11 and Figure 12 present the results of our map building method. In the Figure 11, the map is obtained after 8 acquisitions, the Figure 12 shows the result of this algorithm in a corridor after 20 acquisitions (for more visibility we have removed the segment between points). We can note a few drift in this second example, because the trajectory of the robot is linear, and the environment is symmetric and repetitive.

Conclusion

We have developed a simultaneous localization and map building system based on a cooperation between a stereoscopic omnidirectional perception system and a dead-reckoning system. We have solved the preponderant problem residing in the robust sensorial model construction, using two exteroceptive conical sensors. The optimization of the robustness is obtained with the fusion of complementary treatments. The method allows us to merge several heterogeneous discriminate criteria. This approach takes into account the notion of weighting for each elementary treatment. We have developed a robust absolute localization algorithm based on the matching of the stereoscopic sensorial primitives with the environment map. We have integrated in the matching stage a coherence position test linked to the dead-reckoning estimation, which permits to increase the precision and the robustness of the robot's configuration estimation (a mean accuracy of 10 cm for the position). The matching stage based on the Dempster-Shafer theory allows to estimate the robot's pose and to merge the sensorial primitives with those included in the map. The use of Dempster-Shafer theory and the recursive least squares method with the previous localisation algorithm gives us a coherent construction of the environment field. We can note that our localization and map building paradigm generates a few drift, even on long path. On several stereoscopic acquisitions made in an indoor environment, we obtain a coherent incremental map and an important precision on the considered primitives.

References

[1] Y. Yagi, Y. Nishizawa, M. Yachida, "Map-based navigation for a mobile robot with omnidirectional image sensor COPIS", IEEE Trans. On Robotics and Automation Vol. 11, pp. 634-648, October 1995.

[2] J. Guivant, E. Nebot, S. Baiker, "Autonomous navigation and map building using laser range sensors in outdoor applications", Journal of Robotic Systems, Vol 17, n° 10, pp 565-283, October 2000.

[3] L. Delahoche, C. Pégard, M. Mouaddib, P. Vasseur, "Incremental map building for mobile robot navigation in an indoor environment", Proceedings of International Conf. on Robotics and Automation (ICRA'98), Belgium, May 1998.

[4] S. Clark, G. Dissanayake, "Simultaneous Localisation and map building using millimetre wave radar to extract natural features", Proceedings of International Conf. on Robotics and Automation (ICRA'99), pp 1316-1321, Detroit, May 1999.

[5] J.J. Leonard, H.F. Durrant-Whyte, "Dynamic map building for an autonomous mobile robot", The International Journal of Robotics Research, Vol. 11 n°4, August 1992

[6] J.L. Crowley, "World modelling and position estimation for a mobile robot using ultrasonic ranging", Proc. of Int. Conf. on Robotics and Automation, pp. 674-680, May 1989

[7] H.P. Moravec, « Sensor fusion in certainty grids for mobile robots », AI Magazine, Summer 1988

[8] A. Elfes, "Sonar-based real world mapping and navigation", IEEE Journal of robotics and automation, Vol. RA-3, N°3, pp. 249-265, June 1987

[9] J. Boreinstein, Y. Koren, "Histogrammic in-motion mapping for mobile robot obstacle avoidance", IEEE Trans. on robotics and automation, Vol. 7, N°4, pp. 1688-1693, August 1991

[10] D. Fox, W. Burgard, S. Thrun, "Probabilistic Methods for Mobile Robot Mapping", Proc. of the IJCAI-99 Workshop on Adaptive Spatial Representations of Dynamic Environments, 1999.

[11] B. Kuipers, Y.T. Byun, « A robot exploration and mapping strategy based on a semantic hierarchy of spatial representations », Robotics and Autonomous Systems, 8 1991.

[12] H. Choset and J. Burdick, "Sensor Based Planning, Part I: The Generalized Voronoi Graph," Proceedings of the 1995 IEEE International Conference on Robotics and Automation (ICRA '95), Vol. 2, May, 1995, pp. 1649 - 1655.

[13] C. Drocourt, L. Delahoche, C. Pégard, C. Cauchois, « Localization method based on omnidirectional stereoscopic vision and dead-reckoning », Proc. of the IEEE Int. Conf. on Intelligent Robots and Systems (IROS'99), Korea, pp. 960-965, october 1999.

[14] D. Gruyer, V. Berge-Cherfaoui, « Matching and decision for vehicle tracking in rad situation », Proc. of the IEEE Int. Conf. on Intelligent Robots and Systems (IROS'99), Korea, october 1999.

[15] G.A. SHAFER "A mathematical theory of evidence", Princeton : university press, 1976

[16] D. P. Huttenlocher, G. A. Klanderman, W. J. Rucklidge "Comparing Images Using the Hausdorff Distance", IEEE Transactions on Pattern Analysis and Machine Intelligence, 15(9):850-863, 1993.

Acknowledgement

This work was supported in part by "Région Picardie" under the project "Pôle DIVA" (SAAC project 00-2 - "Système Actif d'Aide à la Conduite").