

# Particle-based Sensor Modeling for 3D-Vision SLAM

Daniele Marzorati\*, Matteo Matteucci†, Domenico G. Sorrenti\*

\*Università di Milano - Bicocca, Milano, Italy

†Politecnico di Milano, Milano, Italy

**Abstract**—Self localization and mapping with vision is still an open research field. Since redundancy in the sensing suite is too expensive for consumer-level robots, we base on vision as the main sensing system for SLAM. We approach the problem with 3D data from a trinocular vision system. Past experience shows that problems arise as a consequence of inaccurate modeling of uncertainties; interestingly enough, we found that accuracy in modeling the robot pose uncertainty is much less relevant than for the uncertainty on the sensed data. To overcome the severe limitation of linear and Gaussian approximations, we applied a particle-based description of the inherently non-normal probability density distribution of the sensed data; the aim is to increase the success rate of data association, which we see as the most important problem. The increase in correct data associations reduces the uncertainty in the model and, consequently, in the robot pose, respectively estimated with a hierarchical map decomposition and a six degree of freedom extended Kalman filter. In this paper, we present approaches for particle-based sensor modeling and data association, with a comparative experimental evaluation on real 3D vision data.

## I. INTRODUCTION

In many real-world applications, e.g. autonomous world model building by a mobile robot (a.k.a. the SLAM problem), modeling the uncertainty of the data is of uttermost importance, as it heavily affects the system performance; see e.g. [1] for applications that were successful because of the appropriate modeling of uncertainties. In this problem, the observer mobility implies the integration of data that, for its nature, is very noisy. Every perception system suffers for its own, usually non-gaussian, noise; the gaussian/normal distribution is often used to model it and often this turns out to be a limitation.

In the SLAM problem, it is nowadays usual ([1], [2]) to model the non-normal robot pose with the particle approach. This has been proven successful in robots based on Laser Range Finders as the main sensing system. In many mobile robots applications (outdoor navigation, indoor navigation with ramps and/or obstacles not perceivable at the LRF height, door opening with door handle, etc.) such sensing system cannot be the only complex sensing system of the robot. On the other hand, a vision-based sensing system, which could leverage on the huge amount of computer vision algorithms, could be the main sensing system of a mobile robot. This is also pushed by economical constraints: *“extensive market analyses show that a complex sensing system for a mobile robot cannot cost more than 10US\$, for a consumer-level robot”* [3]. Today a multi-camera system still does not match such target on cost, but the difference,

w.r.t 3D LRFs, is impressively large in favor of multi-camera systems; moreover, it is fast-reducing.

In this work we show that, for vision-based SLAM, the accurate modeling of the uncertainty on the sensed data allows to reach excellent results, without any particular attention to the uncertainty on the robot pose. Section II briefly introduces the state of the art in SLAM while the issues in uncertainty modeling of 3D data from a stereo-vision systems are discussed in Section III. Section IV describes a novel approach to data association exploiting a particle-based description of data uncertainty. The effectiveness of such idea is presented, on real data, in Section V. Conclusions and future work are presented in the last section.

## II. STATE OF THE ART

A non-linear function of normally distributed variables is usually approximated with a gaussian model; this dates back to the work of [4] and has been widely used in the literature. EKF can be seen as a similar approach, as it also bases on linearization. EKF is also state-of-the-art in SLAM. In the recent years we could appreciate the affirmation of other Bayesian information processing, from Unscented Kalman Filter [5] and (Sparse Extended) Information Filter [6], to Particle Filter [7]. The state of the art in uncertainty modeling for SLAM is to use the normal distribution for the sensed data and a more accurate, particle-based, approximation for the robot pose [2]. For each value of the robot pose a reconstructed world is computed, basing on a normal distribution of the sensed data, given that pose. In these approaches we have a gaussian distribution, to approximate the noise on the observations of the features, and a particle filter to model the robot pose. In this way the different hypotheses on data-association (and therefore on world reconstruction) can be kept under consideration as long as they could turn to be correct (a not greedy approach). In grid-based approaches we have a discretization of the state, the robot location, into cells (histogram filter) and a gaussian distribution is used to model the features [8] This filter represents the cumulative posterior, for each region, with a single probability value.

## III. 3D SENSOR DATA FROM VISION

The perception system we used follows a consolidated approach, i.e. trinocular stereo, and reconstructs the scene in terms of 3D segments [9]. This makes the robot able to capture part of the 3D nature of the environment. Such systems are quite widespread in the computer vision and robotics communities. In order to give out 3D segments

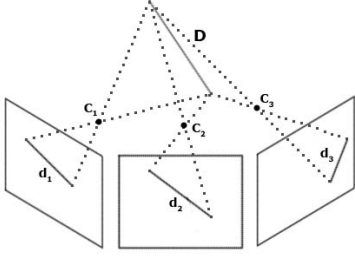


Fig. 1. 3D segment-based reconstruction for a trinocular stereo-system.

the system has to deal with segments since the very first (image) processing steps. The segments are represented by the 3D coordinates of their extrema. The trinocular approach exploits the trinocular epipolar constraint to speed up the search for triplets corresponding 2D segments in the 3 images of the system. Our implementation differs from the original only in the use of the Fast Line Finder [10], in the polygonal approximation phase.

The last step, the most interesting for this work, is the computation of the parameters of the 3D segment, represented by the 3D coordinates of the endpoints. In Fig. 1,  $\mathbf{D}$  is the 3D scene segment,  $\mathbf{C}_i$  and  $\mathbf{d}_i$  are respectively the projection center and the projection of  $\mathbf{D}$  on image  $i$ . The cameras are calibrated with a standard DLT technique, which models the uncertainty on the camera projection matrix as normally distributed. The 2D segments are also measured in the image space as normally distributed. The system therefore produces the 3D extrema of each scene segments as well as an associated covariance matrix, intended to represent the measurement uncertainty of the 3D data as a normal probability distribution. As making evident the limitations of modeling the data uncertainty with a normal distribution is the aim of this work, we developed a particle approximation for the uncertainty of the 3D segments. The normal distributions (expected value and covariance matrix) of the independent variables, i.e. the projection matrixes and the 2D segments on the image plane, are used to generate the samples to which we apply the triangulation process.

3D reconstruction systems based on stereo-vision generate data affected by uncertainties that strongly depend on the observer-to-feature distance. This is due to the triangulation relationship. This uncertainty is relevant in order to correctly associate different measures of the same world feature. Let's consider, for the sake of simplicity, the binocular stereo-system in the plane as in Fig. 2; two cameras are placed at known positions (baseline distance =  $b$ ). A point  $\mathbf{N}$  projects in the images; we measure its projections at  $x_L^N$  and  $x_R^N$ .  $x_L^N$  and  $x_R^N$  are affected by noise, due to many sources, e.g. space and intensity image quantization, etc. The noise on  $x_L^N$  and  $x_R^N$ , in turn, cause the location of  $\mathbf{N}$  to be reconstructed with error. The uncertainty induced by the triangulation, i.e. the intersection of the two interpretation lines of  $x_L^N$  and  $x_R^N$ , is not a simple scalar function of the distance to  $\mathbf{N}$ , it is also skewed and oriented. Nearby

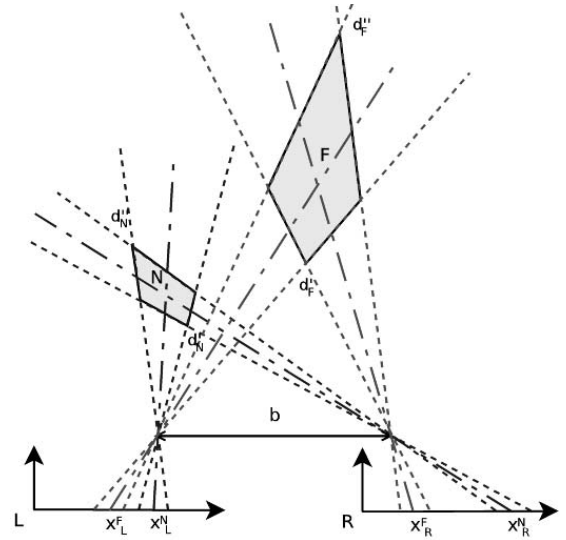


Fig. 2. Uncertainty diamonds for points at different distances

(to the observer) points are affected by a fairly compact uncertainty distribution, whereas farther points are affected by an elongated uncertainty distribution, that is roughly aligned with the line of sight to the point. We want to evaluate and represent this uncertainty, in order to take it into account in any further reasoning, e.g. in a SLAM system.

#### A. Gaussian Modeling

The image coordinates and the projection parameters are here considered as corrupted by a normally distributed noise. The parameters of the distribution have to be determined. As triangulation is a non-linear operation, the distribution is non-gaussian, but in the most used approach it is approximated to a normal. This makes much easier dealing with the errors in the subsequent phases, e.g. data association and integration of the different measures. This approach dates back to [4] and to the work of O. Faugeras's group at INRIA.

Let  $\mathbf{L} = [x_L, y_L]^T$  and  $\mathbf{R} = [x_R, y_R]^T$  be the projections in the left and right images, respectively. Their coordinates are normally distributed random vectors with mean  $\mu_L$  and  $\mu_R$ , and covariance matrices  $\mathbf{V}_L$  and  $\mathbf{V}_R$ , respectively

$$\mathbf{N} = \mathbf{f}^{tri}(\mathbf{L}, \mathbf{R}) = \begin{cases} X = b(x_L + x_R)/(2(x_L - x_R)) \\ Y = b(y_L + y_R)/(2(x_L - x_R)) \\ Z = b/(x_L - x_R) \end{cases} \quad (1)$$

where  $b$  is the baseline, and a unit focal length is assumed.

Although  $\mathbf{f}^{tri}$  is nonlinear the approach stipulates that  $\mathbf{N}$  can be approximated by a normal distribution with average  $\bar{\mathbf{N}} = \mathbf{f}^{tri}(\mathbf{L}, \mathbf{R})$  and covariance computed basing on  $J$ , the Jacobian of  $\mathbf{f}^{tri}$ .

$$\mathbf{V}_N = J|_{\mathbf{L}, \mathbf{R}} \begin{bmatrix} V_i & 0 \\ 0 & V_r \end{bmatrix} J|_{\mathbf{L}, \mathbf{R}}^T \quad (2)$$

#### B. Particle-based Modeling

We sample the normally distributed independent variables (projection parameters and 2D features) with  $n$  particles. We

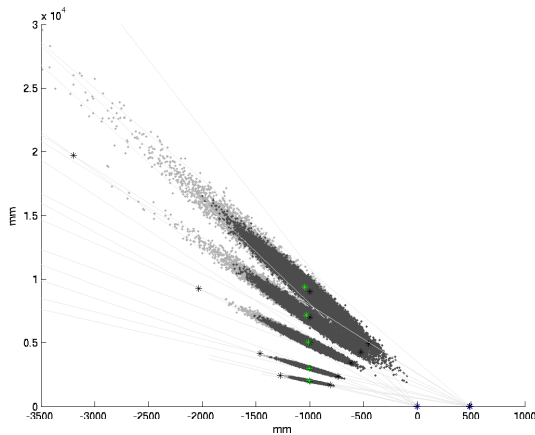


Fig. 3. Shape of the distributions (normal in dark and the unknown in light grey) w.r.t the position of  $N$ ; baseline is from 0 to 500

then propagate the (2D) particles through the triangulation process as if they were 2D image points. We can exploit (1) for each sample. The results are shown in Fig. 3. Note that the differences, between the gaussian approximations and the sampled ones, increase as  $N$  moves away from the observer.

#### IV. DATA ASSOCIATION

The key point of the SLAM process, from a robustness perspective, is *data association*. The goal of such activity is to map uncertain information coming from sensors to uncertain information in the map. In our case the data association algorithm has to map 3D segments in the current view with 3D segments in the current submap. A correct association is obtained when the algorithm is capable to map a pair of segments (one in the view and one in the map) corresponding to the same real segment, perceived from different points of view.

The two different approaches in uncertainty modeling (i.e., gaussian approximation and our proposal of particle-based modeling), can be applied. We claim our proposal increases the data-association success rate. In both approaches the association is based on three geometric criteria, for finding matches between view and map. These criteria use a Mahalanobis distance to evaluate the matches, and has been described with more detail in [11].

##### A. Gaussian Data Association Method

In the following we introduce a simple formalization of the data association, to be used in the rest of the paper. Suppose we measured a set of features  $\mathcal{V} = \{v_i\}$  from a certain robot position. Each feature is represented by its coordinates and its uncertainty. We have also the set of features  $\mathcal{M} = \{m_j\}$  observed and integrated in the map up to that moment. The purpose of a data association algorithm is to find a proper hypothesis of correspondences between them (list of pairs  $v_i$  and  $m_j$ )  $H_k = \{ \langle v_1, m_{j_1} \rangle, \langle v_2, m_{j_2} \rangle, \dots, \langle v_\nu, m_{j_\nu} \rangle \}$ . To generate this hypothesis we must explore an interpretation tree, where each node represents a pair of corresponding features  $\langle v_i, m_j \rangle$ . In [11] we discussed

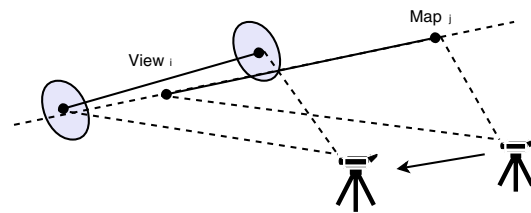


Fig. 4. The “moving window” problem.

some algorithms for data association and the importance of using proper criteria to match features in the view with features in the map. Usually the point-to-point (extremum-to-extremum) distance is considered as an appropriate criterium for single segment matching and much of the effort is devoted in finding a good association strategy for dealing with the exponential complexity of finding the best match for the whole view. In that paper we showed how better criteria for 3D segment matching result in a better data association, almost independent from the algorithm for interpretation tree traversal (i.e. data association). The approach we proposed is based on a multi-criteria evaluation for associating segments in the View with segments belonging to the Map. The reason for the poor performance of the point-to-point criterium is mainly due to the problem of moving-field-of-view in the sensing system, which turns in a moving window on the world feature, see Fig. 4. The segments extrema could be due to the reduced field of view and not always related to real extrema in the world; this can easily become a problem for the classical point-to-point distance<sup>1</sup>. The  $v_i$  segment is moved into the Map reference frame by means of an estimate of the robot pose, e.g. the one provided by odometry. Such estimate is known altogether with an estimate of its uncertainty, which is propagated to the segment in the Map reference frame. The data association criteria are based on uncertain Mahalanobis distance.

We propose three criteria to validate an hypothesis of match between  $v_i$  and  $m_j$ :

- 1) 1<sup>st</sup> criterium: Mahalanobis distance between the support line of the map segment and the extrema of the view segment;
- 2) 2<sup>nd</sup> criterium: Angle between the support lines of the two segments: this constraint complements the previous criterium and it is also based on a statistical test that considers the uncertainty in the angle computation;
- 3) 3<sup>rd</sup> criterium: this is not a probabilistic one; it is related to the projections of the segment extrema on the other segment; one of the following two conditions has to hold to get an association:
  - a) at least one of the extrema of the view segment has to project onto the map segment;
  - b) both the extrema of the view segment project outside the map segment, but the extrema of the

<sup>1</sup>Our proposal is of interest also for 2D-3DoF SLAM systems which groups data points into 2D lines, because this moving-field-of-view issue applies there too.

map segment project inside the view segment.

At the end of the process we have a list of possible matches between  $v_i$  and  $m_j$ . We consider only one possible match ( $m_j$ ) for each  $v_i$  segment and use two algorithms to perform this task: *Multi-Criteria Nearest Neighbor* and *Multi-Criteria Joint Compatibility*.

Multi-Criteria Nearest Neighbor is a variation on the nearest neighbor data association algorithm that bases on the classic ICP (Iterative Closest Point) algorithm [12] and the multi-criteria approach. Moreover, it orders the list of all potential matches, after each association decision. We call this approach *Full-Order* (FO), and the whole NNFO.

The Multi-Criteria Joint Compatibility (JCBB) integrates the three criteria in the joint compatibility branch and bound technique [13]. To reduce the computational time required to explore one branch of the interpretation tree, we first compute the  $3^{rd}$  criterium on the couple  $\langle v_i, m_j \rangle$  because it is very selective and fast. We then compute the  $2^{nd}$  criterium on the same couple, to decrease the possibility to compute in vain the next step. Finally, we use the  $1^{st}$  criterium, in the joint compatibility test.

### B. Particle-based Association Method

Our claim in this paper is that a further improvement in data association can be obtained using a proper uncertainty modeling; to support such claim, we modified the data association algorithm to consider a particle-based model of the uncertainty in the sensed data. Particles represent hypothesis for the real position of 3D extrema of reconstructed segments; each particle, in this approach, represents a single possible extremum of the segment and it is known precisely.

We sample a particle from the two distributions, of the two extrema. This pair of samples is the  $v_i^k$  sample for segment  $v_i$ . We move then  $v_i^k$  into the map reference frame (as for the previous method) by means of an estimate of the robot pose, e.g. odometry. The uncertainty on the particle depends only on the error on the robot pose, which we model with a gaussian distribution. This uncertainty is computed with a classic error propagation. This approach contrasts state-of-the-art SLAM methods, that use particles to model pose uncertainty.

We now apply the three criteria to each particle-segment  $v_i^k$ , w.r.t. all the map segments. If  $v_i^k$  passes the tests with a map segment  $m_j$ , then a specific counter for  $m_j$  is incremented. If the number of particle-segments  $v_i^k$  for some  $m_j$  is larger than a threshold, then the match  $\langle v_i, m_j \rangle$  is admissible. The process is repeated for all view segments. This approach can be seen as the empirical computation of the matching probability integral using particles distribution. At the end a full-order algorithm is executed to extract a list of valid matches:

- 1) for all  $m_j$  from  $\mathcal{M}_j$  and  $v_i$  from  $\mathcal{V}_i$ 
  - a) sample a set of  $K$  particles for each extrema of  $v_i$  and move these segments  $v_i^k$ s into the Map reference frame
  - b) compute the number of  $v_i^k$  that pass the 3 tests; this count is then compared with a threshold ( $\alpha$ ),

relative to the total number of particles for  $v_i$ ; if larger, then  $v_i$  can be associated to  $m_j$

- 2) order the matches  $\langle v_i, m_j \rangle$  in decreasing order
- 3) select as match the first pair  $\langle v_i, m_j \rangle$
- 4) erase any other pair that involves  $v_i$  or  $m_j$  from the list, and order the list
- 5) repeat from the selection of the best match until no match is possible

The method (PFO) is therefore depending on  $n$ , the number of particles, and the threshold  $\alpha$  (PFO $n\alpha$ )

## V. EXPERIMENTAL ACTIVITY

For the experimental activity we used a mobile robot from Robosoft, which computes odometry as a 3DoF pose  $[x, y, \theta]^T$ ; this datum reaches a PC via serial line. On the PC we have an Eltec frame grabber capable to grab three 704x558x8 pixels images at the same time. Each channel of the frame grabber is connected to a Sony XC75CE camera. Cameras have been calibrated with a standard DLT approach. The robot has been moved inside the 4th floor of building U7, Univ. Milano - Bicocca, Milano, Italy. The robot travelled about 100m, activating its perception system every few centimeters, for a total of 1000 views (activations), with about 15 segments per view.

To generate the 3D particles (segments extrema) we sampled 100 or 1000 (depending on the experiment) particles from the normal uncertainty of each 2D extremum. This is repeated for the 6 extrema (3 segments, 1 per image); we also sample the projection matrix from the 11-variate normal given out by our calibration tool. Next, we compute the 3D segment extrema with the triangulation process. This procedure is repeated 100 or 1000 times, for each extremum of each 3D segment. Subsequently, we use this information to reconstruct the scene with a SLAM process. We need to clarify what happens currently, when a view segment  $v_i$  is associated to a map segment  $m_j$ , in order to integrate, i.e. fuse, the 2 measures. If both  $v_i$  and  $m_j$  are modeled with normal distributions, then the usual fusion applies. On the other hand, when  $v_i$  is represented by a particle distribution, we use the first two moments of it to update the normal distribution of  $m_j$ .

Table I shows the comparison between the different data-association methods. The leftmost (NNFO) is a frequently used approach, based on normal distributions to describe the extrema of the segments. The second is the state-of-the-art JCBB, again with normal distributions. The others are based on the new method, with distributions approximated by particles. PFO100 stands for 100 particles per extremum, PFO1000 for 1000 particles. The percentage is the threshold mentioned before ( $\alpha$ ), i.e. the minimum number of particles, w.r.t. the total number of  $v_i^k$ , which have to match with  $m_j$ , in order to associate  $v_i$  to  $m_j$ .

Being un-available the ground truth of the explored environment, we performed the comparison looking at the number of matches generated during the view-integration phase, before any loop-closure. Also the number of submaps,

TABLE I  
1000 VIEWS, NO LOOP CLOSURE

.	NNFO (norm)	JCBB (norm)	PFO100 50%	PFO100 10%	PFO1000 50%	PFO1000 10%
# matches	1153	1338	1545	1585	1545	1599
# submaps	15	11	7	6	7	6
time (min)	30	62	201	189	1740	1980

created during this exploration, is reported because it contributes to shed light on the amount of correct associations performed by each approach.

It has to be stressed that by not caring of data-association problems the overall map results into larger number of false segments, all apparently independent, but actually referring to the same scene features. On the other hand, approaching the issue by simply increasing the normally modeled uncertainty turns into a large number of wrong associations.

We are very excited about the results obtained with our proposal and do not think the processing time to be really relevant at the moment, given that current techniques were indeed unable to reach the goal of building a good map. Nevertheless, we reported the time taken for the processing on a P4 laptop, starting from images on files, and including stereo-matching, just for giving an idea on this aspect of our preliminary results. We also note that particle approaches can be easily and effectively implemented on parallel architectures.

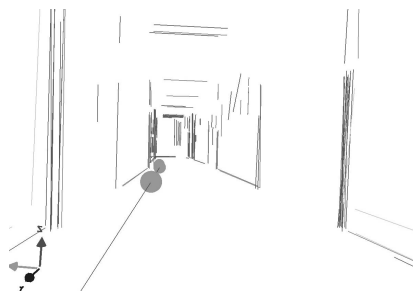


Fig. 5. A view of the map obtained with normal modeling of uncertainty

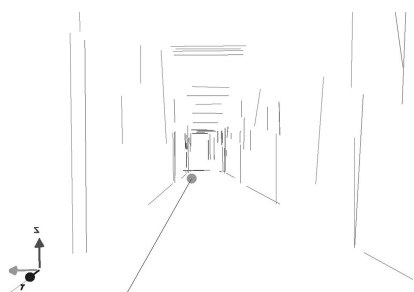


Fig. 6. A view of the map obtained with particle modeling of uncertainty

By looking at the table we can observe that our proposal allows about 34% more matches w.r.t. NNFO (norm) and about 14% w.r.t. JCBB (norm). If looking at the number of submaps (size limit was set at 50 features for all approaches) the reduction appears even larger: about 26% w.r.t. NNFO (norm) and about 36% w.r.t. JCBB (norm).



Fig. 9. Two reconstructed scene segments

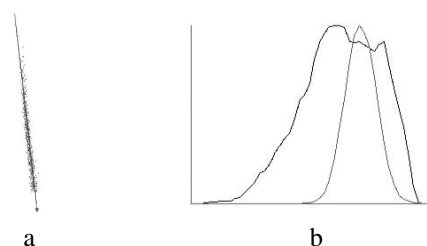


Fig. 10. Segment 1, extremum 2; see text for details.

The effectiveness of the approach can be further appreciated by looking at Fig. 5 and 6; they show a similar view on the NNFO (norm) and PFO1000 10%, respectively, 3D reconstructions. Notice the large reduction in the number of un-matched segments when going from the normal to the particle approach. A view from the top of the two maps is presented in Fig. 7 and 8.

In order to support our claim about the need for accurate modeling of uncertainties in the sensed data we present a couple of real extrema; in Fig. 9 notice 2 scene segments: the green endpoints have coordinates  $[2.9, 8.6, 2.8]$ ,  $[1.6, 9.2, 2.9]$ , while the extrema of the red are  $[1.4, 9.0, 2.8]$ ,  $[0.8, 9.0, 2.8]$ . Fig. 10a shows the uncertainty (in the  $x, y$  plane) for an extremum of the green segment, using the particle model, while the Fig. 10b shows the comparison between this uncertainty model and the uncertainty computed using the Jacobian method, along the line connecting the extrema of the uncertainty diamond. The direction of the arrow, in the left figure, corresponds to direction of the abscissa in the figure on the right. Fig. 11a and b presents the same for the first extrema of the red segment. Notice the differences between the shape of the normal and the sampled distributions.

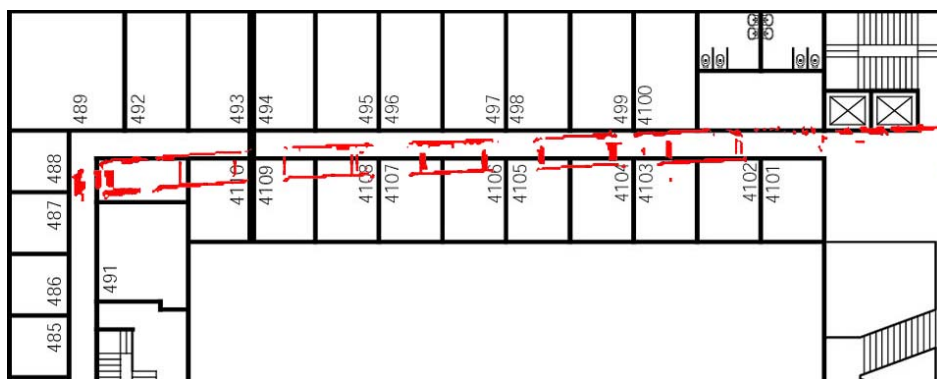


Fig. 7. A view from the top of the map (normal modeling of uncertainty)

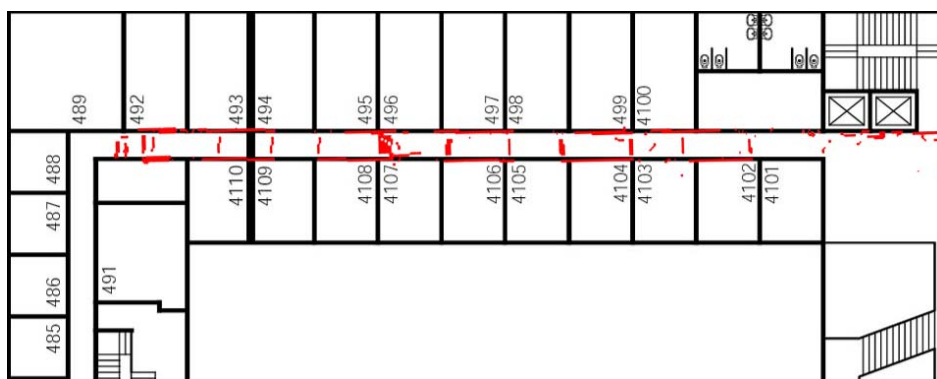


Fig. 8. A view from the top of the map (particle modeling of uncertainty)

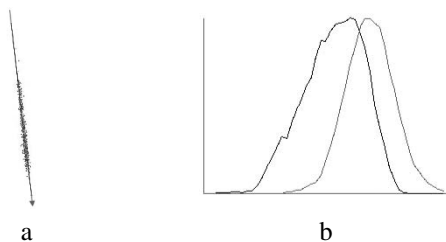


Fig. 11. Segment 2, extremum 1; see text for details.

## VI. CONCLUSIONS

In this paper we presented a particle-based approach to uncertainty modeling of 3D data sensed through a stereo vision system. Our claim in doing this is that it is possible to obtain a better data association exploiting such modeling, with respect to the classical gaussian approximation. Impressive results have been obtained on real data, confirming that a proper uncertainty model is a key factor in data association and in SLAM in general. It is worth noticing that this approach apparently contrasts with the state of the art in SLAM algorithms, which usually base on a non gaussian (particle-based) estimate of the robot pose, while perceptions are treated as normally distributed.

We are currently working to reduce the time required to compute a particle-based representation and to improve the speed of the method.

## REFERENCES

- [1] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. MIT Press, 2005.
- [2] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit, "Fastslam 2.0: An improved particles filtering algorithm for simultaneous localization and mapping that provably converges," in *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, 2003.
- [3] C. Angle, "Inv. talk at euron meeting," Amsterdam, March 2004, the author is CEO of iRobot inc.
- [4] L. Matthies and S. Shafer, "Error modeling in stereo navigation," *IEEE Journal of Robotics and Automation*, vol. RA-3, no. 3, pp. 239–248, June 1987.
- [5] S. Julier and J. Uhlmann, "A new extension of the kalman filter to non-linear system," in *Proceedings of International Symposium on Aerospace/Defence Sensing, Simulate and Controls*, 1997.
- [6] P. Newman and H. Durrant-Whyte, "A new solution to the simultaneous localization and map building (slam) problem," in *Proceedings of SPIE*, 2001.
- [7] D. Blackwell, "Conditional expectation and unbiased sequential estimation," *Annals of Mathematical Statistic*, vol. 18, pp. 105–110, 1947.
- [8] J. Guivant, E. Nebot, J. Nieto, and F. Masson, "Navigation and mapping in large unstructured environments," *International Journal of Robotics Research*, vol. 23, 2004.
- [9] N. Ayache, *Artificial vision for mobile robots*. MIT Press, 1991.
- [10] P. Kahn, L. Kitchen, and E. Riseman, "A fast line finder for vision-guided robot navigation," *IEEE Transaction on PAMI*, vol. 12, no. 11, November 1990.
- [11] D. Marzorati, D. G. Sorrenti, and M. Matteucci, "Multi-criteria data association in 3d-6dof hierarchical slam with 3d segments," in *Proc. of ASER06 Conf.*, 2006.
- [12] P. J. Besl and N. D. McKay, "A method for registration of 3-d shapes," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 239–256, Feb. 1992.
- [13] J. Neira and J. D. Tardós, "Data association in stochastic mapping using the joint compatibility test," *IEEE Trans. R&A*, vol. 17, no. 6, pp. 890–897, 2001.