3D-6DoF Hierarchical SLAM with 3D vision

D. Marzorati[1], M. Matteucci[2], and D. G. Sorrenti[1]

[1] Universitá di Milano - Bicocca, DISCo, Milano, Italy,
{marzorati,sorrenti}@disco.unimib.it,
[2] Politecnico di Milano, DEI, Milano, Italy
matteucci@elet.polimi.it

Abstract. We motivate and present a SLAM system capable to deal with data from 3D segment-based vision system. These are widespread systems in robotics. Reliable world mapping in large indoor environments is demonstrated by the experimental activity.

1 Introduction

Simultaneous Localization and Mapping (SLAM) deals with the automatic construction of a geometric model of environment [1]. The main issues are related to errors in robot localization and in mapping the world; as a result the world model is affected by geometric inconsistencies. The absence of a reliable SLAM functionality prevents practical use of mobile robotics technology whenever an a priori and up-to-date map of the workspace is not available, i.e. nearly always, as executive drawings (if available) differ from reality, day-to-day usage of space introduces changes such as un-fixed furniture, temporary obstacles, etc.

Many approaches are known in the literature of SLAM systems; some of the most known are Fastslam [2], which decompose the problem in two: robot localization and estimation of the position of the world features and then makes use of a modified particle filter for the estimate of the robot pose and EKF for the map. Graphical SLAM [3] represents the world map as a graph where nodes carry information about the pose of each world feature and the robot; on the edges are the relationship between nodes. Many other approaches base on EKF for building a geometrically consistent map. An interesting technique is Hierarchical SLAM [4], which is based on EKF, Mahalanobis distance, interpretation trees, and hierarchical decomposition of the data structure (which allows a reduction in complexity by limiting the items involved in the most cumbersome SLAM phases. This approach has been used with many sensing systems (sonar, laser range finders, etc.) and also with a 3D vision system like the one we used, but with data obtained by projecting on the floor the 3D data.

In this paper we first shortly recall in Section 2 the specific aspects of our sensed data, and then introduce (Section 3) our 3D-6DoF Hierarchical SLAM work. We then illustrate in Section 4 the experimental activity performed.

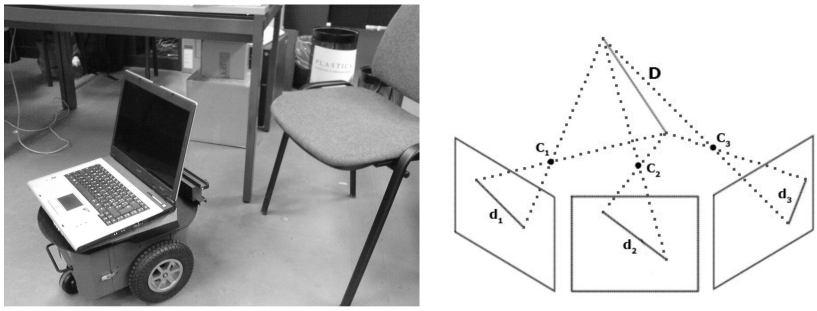
2 Sensor Data

Some robot activity requires a full 3D knowledge of the observed environment features; a few examples are: tables legs or steps which constrain motion; door handles and fire extinguishers help in localization; cleaning has to be performed under tables and chairs; fire extinguishers have to be avoided; books to be moved are on top of tables, etc. In Figure 1a the robot can navigate/clean under the table, but not under the seat. It is therefore relevant to map the real 3D robot workspace, but most of these items are not 3D-perceivable with 2D polar map sensors like LRFs or omnidir. vision. Some existing work dealing with 3D data bases on 3D LRFs [5], but these devices provide just geometric data, which makes difficult other robot tasks. An example could be the semantic classification of places [6], which are required for a real indoor service robot. Even though we are here proposing to use just the geometry provided by 3D vision systems (because we are working with the geometric task of map building), we think that the full richness of vision systems output is necessary for other tasks. On the other hand, it is difficult to put many sensing systems on the same robot, typically because of the cost limitations, like in consumer-level robotics. These considerations are our main motivations for a vision-only 3D-data-based SLAM approach.

The sensing system we use gives out 3D segments, and it is based on the trinocular approach [7]. It deals with segments since the very first processing step. Hence it looks for 2D segments in the image, and then for correspondences between the different images. The last step is the computation of the parameters of the 3D segment, represented by the 3D coordinates of their endpoints. In Figure 1b, $\mathbf{D}$ 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. Cameras are calibrated altogether with their covariance matrix so that 3D segment endpoints can be given out altogether with an associated covariance matrix, to represent the measurement uncertainty as a normal probability distribution. Such systems date a long time ago and are quite widespread in the computer vision and robotics communities. Our implementation differs from the original only in the use of the Fast Line Finder [8] in the polygonal approximation phase.

3 3D-6DoF Hierarchical SLAM

In the notation kD-lDoF SLAM, the first item (k) refers to the dimensionality of the data used for building the world model. We use 3D data from the perception system mentioned before. The second item (l) refers to the dimensionality of the observer pose. We model the pose as a full rigid-body transformation, i.e. a 6DoF pose. The whole system is a 3D-6DoF SLAM system, like the one in [5]. On the other hand, the system in [9], which is also based on data from a trinocular system, because of the projecting of data on the floor, is a 2D-3DoF SLAM system. This is a quite common approach in indoor mobile robotics, where the robot is moving on a supposedly flat floor, and it looks reasonable to represent the robot pose with a $< x, y, \vartheta >$ triplet. Our past experience shows that such



(a) (b)

Fig. 1. (a) Cleaning under table/seat requires a 3D understanding of the free space. (b) 3D segment-based reconstruction for a trinocular stereoscopic system.

flatness hypothesis is not realistic and that even with small floor imperfections a complete pose representation, i.e. 6DoF, is of great advantage even in indoor SLAM. In previous work we showed that a higher accuracy can be obtained in pose recovery from images of a scene if a complete 6DoF model of the pose is used [10]. We showed also [11] that a complete, i.e. more realistic, modelling of uncertainties turns also into higher accuracies. In other words, un-modelled uncertainties, i.e. use of deterministic values just due to modelling laziness, bias the estimates. The application scenario in [11] was object localization. We are therefore claiming that a complete (and realistic) modelling of the reality, i.e. 6DoF instead of 3DoF for the robot pose, is of great advantage even in an indoor SLAM scenario.

3.1 Views and Submaps

A view is the set of 3D segments given out in one activation of the perception system. Each segment endpoint is a triple of coordinates, its uncertainty is a 6×6 covariance matrix. Each local map, or submap, is the result of the combination of some views. It contains an estimate of each world feature, i.e. segment, as the result of possibly many observations of it, altogether with the estimate of the robot pose. These data are referred to the same reference frame, which is called "base reference" of the submap. Each submap is therefore a representation of a part of the environment. An important property of submaps is the stochastic independence with respect to other submaps.

At first a new submap is initiated with the output of the perception system. The base reference is put on the current robot pose. The uncertainty on the base reference is null [12]. After some motion the perception system is activated again and a new view generated. The integration of views processing, for details see [13], combines the data in the new view with the data in the submap, e.g. creating a single instance of a world feature from the (possibly) two measures (submap and view) available. Odometry plays a role here because it gives an estimate of the motion. The processing is based on EKF; the state is the union of the vectors of the features and the 6DoF robot pose. Associations are used to update the state. The robot pose, being part of the state, is updated according to the new view data; on long motion sequences this helps limiting to some extent the cumulative odometric errors, see Section 4.

This process is repeated as long as some termination criteria are false. Then the submap is closed and a new submap is initiated at the current robot pose.

3.2 Global Map and Loop Closure

The global map maintains the relationships between submaps as an oriented graph; submaps are the nodes, while the edges represent the spatial relationships between the base references of two submaps, which are therefore considered independent in the features. When closing a submap (i), the last robot pose ($\mathbf{x}$) becomes the base reference of the new submap (j). This pose ($\mathbf{x}_j^i$), with respect to the closed submap base reference, is stored in the graph edge connecting the two submaps.

The loop closure process is a complex activity that involves many steps. The first is obviously *Loop Detection*, i.e. detecting a submap close to the one just closed, that is involved in a loop closure. Once a loop has been detected it is necessary to perform *Data Association* to extract common features. This is obtained by a procedure that seeks an hypothesis H that connects each feature in the first submap to the (possibly) corresponding feature in the second submap. This hypothesis is used to determine both robot and features poses in the submaps. There are different approaches to find H; we use an interpretation tree, as done in [4], exploiting an adapted RANSAC algorithm, which bases on a version of the joint compatibility test [9], adapted to the 3D-6DoF problem. Once data-associations have been found, it is possible to perform *Robot Relocation* and *Local Map Joining*, i.e. to estimate the spatial relationship between the two submaps and thus join the two w.r.t. a common reference frame, therefore creating a single submap. The robot pose is changed w.r.t. the new reference frame as well. *Loop Closure* is the final step in global map building, which allows to reduce the errors in the spatial relationships between the submaps in the loop. Link relaxation in loop closure has been re-formulated as a maxima a-posteriori estimate of all base reference poses under the loop constraint $h(\mathbf{x}) = 0$. To solve this minimization problem we used a Sequential Quadratic Programming approach, similar to what done in [4], which is derived from the Kuth-Tucker equations and has been adapted to the 3D-6DoF problem.

4 Experimental Activity

For the experimental activity we used a mobile robot from Robosoft which computes odometry as a 3DoF pose; this datum reaches a PC via serial line. On the PC we have an Eltec frame grabber capable to grab three 704x558x8 pixels

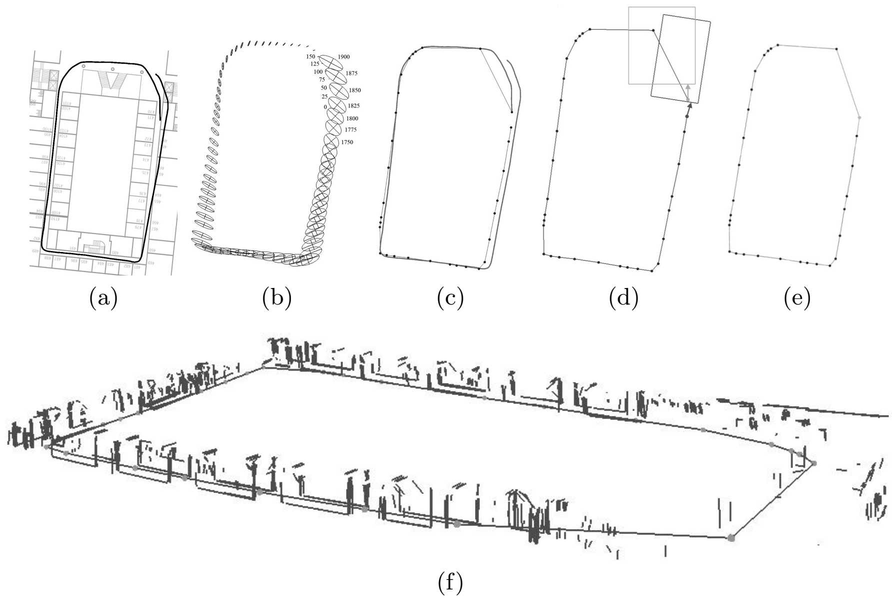


Fig. 2. (a) Odometric travel superimposed to the environment planimetry. (b) Odometric error ellipses ($\pm 3\sigma$), with view_id. (c) Odometric travel (full) superimposed to the base references of the submaps (circles connected by lines); notice the larger accuracy provided by fusion of views. (d) Bounding boxes of last (darker) and first (lighter) submaps. (e) The base references of the submaps after graph relaxation, compare with the base references in (c) or (d). (f) A 3D view of the 6DoF-pose final reconstruction; the solid-circle line is the same as in (e).

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, by hand due to a servo-amplifier failure, inside the 4th floor of building U7, Univ. Milano - Bicocca, Milano, Italy. Distances between consecutive robot poses, i.e. views, were about 0.05m. The overall distance travelled has been about 200m. In Figure 2a the odometric travel is shown, altogether with the environment planimetry. The odometric error is modelled as zero mean Gaussian, and the propagation of this error is shown in Figure 2b with the usual 99% ellipses; notice that the actual, i.e. first, poses are in good agreement with the uncertainty propagated up to the last ones. Submap termination is currently set on the cardinality of the features in the submap; the value used in the reported experiment is 50. In Figure 2c the odometric travel is superimposed to the base references of the submaps, i.e. what could be considered as the overall result of the integration of views processing. This figure shows that the integration of views gives a large increase in accuracy, even though this is not enough for obtaining a geometrically consistent map. When a submap is closed loop-detection is activated; in Figure 2d the bounding boxes of the first and last submaps, i.e. the ones for which a loop is detected, are shown. On these two submaps Robot Relocation and Local Map Joining are applied. At the end the two submaps are fused together in a single submap. The geometric consistency is still not attained at this stage. Loop Closure distribute the errors along the whole set of submaps, i.e. relative poses of submaps. The result of such iterative non-linear optimization (graph relaxation) is shown, in terms of base references, in Figure 2e. A 3D view of the 6DoF-pose final reconstruction is presented in Figure 2f.

5 Conclusions

We presented a SLAM system capable to deal with data from 3D segment-based vision system. These are widespread systems in robotics, but SLAM systems basing on them are not common in the literature. Reliable world mapping in large indoor environments is demonstrated by the experimental activity presented.

References

1. J. D. Tardós and J. A. Castellanos, *Mobile robot localization and map building : a multisensor fusion approach*, Kluwer Academic, 1999.
2. S. Thrun, M. Montemerlo, D. Koller, B. Wegbreit, J. Nieto, and E. Nebot, "Fast-slam: An efficient solution to the simultaneous localization and mapping problem with unknown data association," *J. of Machine Learning Research*, 2004.
3. J. Folkesson and H. I. Christensen, "Graphical slam - a self-correcting map," in *Proc. IEEE ICRA*, April 2004, pp. 383–389.
4. C. Estrada, J. Neira, and J. D. Tardós, "Hierarchical slam: real-time accurate mapping of large environments," *IEEE Trans. on Robotics*, to appear.
5. H. Surmann, A. Nüchter, K. Lingemann, and J. Hertzberg, "6d slam: Preliminary report on closing the loop in six dimensions," in *5th IFAC Symp. IAV*, June 2004.
6. A. Rottmann, O. Martínez Mozos, C. Stachniss, and W. Burgard, "Place classification of indoor environments with mobile robots using boosting," in *Proc. of AAAI*, Pittsburgh, PA, USA, 2005.
7. N. Ayache, *Artificial Vision for Mobile Robots*, The MIT Press, 1991.
8. P. Kahn, L. Kitchen, and E. M. Riseman, "A fast line finder for vision-guided robot navigation," *IEEE Trans. on PAMI*, vol. 12, no. 11, Nov 1990.
9. 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.
10. R. Arpini, V. Caglioti, E. Camnasio, M. Cappelletti, and D. G. Sorrenti, "Visual self-localisation by means of natural landmarks for mobile robot navigation," *Automazione e Strumentazione*, vol. 43, no. 5, pp. 105 – 115, May 1995, in italian.
11. V. Caglioti, F. Mainardi, M. Pilu, and D. G. Sorrenti, "Improving pose estimation using image, sensor and model uncertainty," in *Proc. of IEE BMVC Conf.*, 1994, vol. 2, pp. 805–819.
12. J. A. Castellanos, J. Neira, and J. D. Tardós, "Limits to the consistency of ekf-based slam," in *5th IFAC Symp. IAV*, 2004.
13. D. Marzorati, M. Matteucci, and D. G. Sorrenti, "Multi-criteria data association in 3d-6dof hierarchical slam with 3d segments," in *Proc. of ASER*, 2006.