Proceedings of the 2007 IEEE
International Workshop on Safety, Security and Rescue Robotics
Rome, Italy, September 2007

Building multi-level planar maps integrating LRF, stereo vision and IMU sensors

Luca Iocchi and Stefano Pellegrini and Gian Diego Tipaldi
Dipartimento di Informatica e Sistemistica
Sapienza University of Rome, Italy
{iocchi,pellegrini,tipaldi}@dis.uniroma1.it

Abstract — Building maps of the explored environment during a rescue application is important in order to locate the information acquired through robot sensors. A lot of work has been done on mapping 2D large environments, while the creation of 3D maps is still limited to simple and small environments, due to the costs of 3D sensors and of high computational requirements. In this paper we analyze the problem of building multi-level planar maps. These maps are useful when mapping large indoor environments (e.g., a multi-floor building) and can be effectively created by integrating robustness and efficiency of state-of-the-art 2D SLAM techniques using 2D laser range finder data, with the use of a precise IMU sensor and effective visual odometry techniques based on stereo vision for measuring plane displacements. The main advantages of the proposed solution, with respect to other kinds of 3D maps, are the low-cost of the sensors mounted on the robots and the possibility of exploiting results from 2D SLAM for exploring very large environments. Preliminary experimental results show the effectiveness of the proposed approach.

Keywords: *SLAM, 3D maps, sensor integration*

I. INTRODUCTION

Building 3D models is important in many applications, ranging from virtual visits of historical buildings, to game and entertainment, to risk analysis in partially collapsed buildings. Existing systems for building 3D representation of environments have been developed at different scales: city, buildings, indoor environments, objects, presenting many differences in the sensors and in methods used to acquire data, in the techniques used to process the data, and in the kind of result computed.

Many different sensors have been used for data acquisition. Cameras are the main sensors, since they provide images that contain a very high amount of information: geometry of the scene, colors, textures, etc. However, these data are very difficult to analyze, since Computer Vision problems are still very challenging in real, unstructured environments. To retrieve information about the geometry of the environment, 2D and 3D Laser Range Finders (LRF) are very useful since they provide very precise measurements of the environment. In fact, mapping 2D or 3D environments with LRF has been an active research topic in the last year (this problem is also known as Simultaneous Localization and Mapping (SLAM)) and many systems have been demonstrated to be very effective in this task (specially for 2D environments). However, the use of 3D Laser Scanners is very expensive, while using 2D LRF mounted on pan-tilt unit allows for scanning 3D data, but it requires some time due to the movement of the sensor.

In this paper we propose a solution for a special case of 3D SLAM, that is mapping environments that are formed by multple planar areas (e.g., a multi-floor building). This problem can be effectively solved by decomposing 3D SLAM in two parts: 1D SLAM (to detect the number of planes and cluster sensor readings according to such planes) + 2D SLAM (to build many planar maps). Subsequently, for each pair of adjacent 2D maps, visual odometry techniques are used to determine their displacement, allowing to generate a metric multi-level map of the environment.

The approach described here has been tested on an autonomous robot equipped with a 2D laser range finder, a stereo vision system, and an inertial movement unit.

The paper is organized as follows. Section II describes related work and compares our approach with previous research in this field. Section III presents an overview of the proposed system, while Sections IV describes the proposed solution to multi-level mapping. Section V shows some results of the proposed system, and, finally, Section VI draws some conclusions and presents ideas for future work.

II. RELATED WORK

Several approaches have been presented for 3D environment reconstruction, using different sensors (cameras, stereo cameras, multiple 2D LRF, 3D LRF, and combinations of them). For example, [1] use active stereo vision for building a 3D metric map of the environment, [2], [3] use two orthogonal 2D LRF to build 3D maps of indoor and outdoor environments, while [4] use a 2D LRF mounted on a tilt unit that is able to acquire a very precise 3D scan of the environment with a relative cheap sensor, but it requires a higher acquisition time due to the rotation of the laser. The generation of large 3D maps of portions of a city is considered in [3]; data acquisition is performed through a truck equipped with a horizontal 2D laser scanner (for localization), a wide angle camera and a vertical 2D laser scanner for reconstructing the building's facades. Obstacles, such as trees, cars or pedestrians, are removed considering their relative depth, while holes in the facades arising from the presence of obstacles and from the presence of specular surfaces, are filled through interpolation. The localization was achieved with the help of

978-1-4244-1569-4/07/$25.00 2007 IEEE.

aerial images, thus increasing the cost requirements of such a system. On the other hand, approaches based on feature extraction and computer vision techniques have been proposed (e.g., MonoSLAM [5]), providing for 3D feature-based maps. Outdoor mapping has also been investigated. For example, in [6] the use of stereo vision and visual odometry has been proposed for long distances outdoor navigation of a mobile robot.

Outdoor mapping has also been invistigated: [6] uses stereo vision and visual odometry for long distances outdoor robot navigation, [7] introduces a novel representation for outdoor environment called multi-level surface maps. The world is modelled as a grid, which cells store multiple surfaces. In [8], an actuated laser range finder is used toghether with a camera for outdoor robot mapping. The laser is used to incrementally build a 3D point cloud map of the environment. The images obtained from the camera are used to detect a loop closure events, using an appearance-based retrieval system. While promising, the last two approaches acuire data in a move-stop-move fashion, slowing down the motion of the robot.

All these approaches are focused on building metric or feature based maps, either considering relative small environments to map or focussing on the navigation capabilities of an autonomous platform.

The approach described in this paper aims at combining the robustness and efficiency of 2D SLAM techniques with the requirement of using relatively low-cost sensors and of performing a fast exploration. The main idea is to consider a multi-level planar environment and to perform an off-line analysis of the 3D data, in order to cluster them in many sets each belonging to a single plane. On each of these sets of data coming from a planar environment 2D SLAM techniques are applied and then these sub-maps are merged together using visual odometry techniques. By using state-of-the-art 2D SLAM methods based on laser data and robustness of visual odometry techniques, we can obtain reliable and effective metric multi-level maps of large environments.

III. OVERVIEW OF THE SYSTEM

The system we have developed is based on a mobile robot that carries different sensors for data acquisition. The robot collects and stores data from the environment, these data are processed off-line to build a multi-level map of the environment, possibly including additional information (e.g., snapshots of victims, temperature, etc.) in the map.

The robot used in the experiments is a Pioneer 3 equipped with an on-board PC, a wireless connection to a base station, and 4 sensors: a 2D SICK Laser Range Finder, a Videre Stereo Camera, an XSens IMU, a thermo sensor. The first three are used for mapping and navigation, while stereo vision and thermo sensor are used for victim detection. The robot has software components for autonomous exploration based on an on-line fast mapping approach [9], and thus it can be operated in three modalities: fully autonomous (i.e., the robot runs the autonomous exploration algorithm), partial autonomous (i.e., the user can specify target locations to reach and the robot is able to go there), fully tele-operated (i.e., the robot is controlled by a human operator through a wireless network connection).

For the purposes of the map building process described in this paper, the main goal of the robot is to gather data from the environment, while exploring it, and to store these data on a local disk. The data will be processed off-line at a later stage, possibly on another machine. More specifically, we store all the sensor readings with a 10 Hz frequency (except for stereo images that are acquired at 1 Hz): for each frame, we memorize 180 readings for the 2D LRF, a pair of images from the stereo camera (color left image and disparity image), and 6 values from the XSens IMU. All these data are synchronized with a time-stamp that is generated by the PC on board the robot.

In these experiments, data have been acquired through autonomous exploration, although other modalities would have been adequate too. Further details on the data collected for the experiments are reported in Section V.

In the next section, we will focus on a solution for a special case of 3D-SLAM, where the objective is to create several 2D maps and to establish the displacement among them. This is realized as composition of two process: a) 1D SLAM on the Z coordinate, under the assumption of multi-planar environment this allows for determining the number of planes present in the environment and to associate sensor readings to each detected plane; b) several 2D SLAM processes, one for each detected plan, using the corresponding sensor readings as determined in the phase a).

The relative poses among the different 2D maps are then computed by visual odometry processes that are executed from positions belonging to different and adjacent planes.

IV. 3D SLAM THROUGH 1D + 2D SLAM

In order to create a multi-level planar map, we cannot use only a 2D SLAM algorithm. Yet we do not need to use a full 3D SLAM algorithm. Indeed, we can exploit the assumption that the environment that we wish to reconstruct is piecewise planar. Our idea is that we can still use a robust 2D SLAM algorithm [10] to acquire the map in a planar section of the environment. But, to cope with the transitions between different levels, we use the IMU sensor together with the stereo camera. In particular the IMU is used to detect the transitions, while the visual odometry is used to compute the movement of the robot in this transition phase, where, otherwise, the 2D laser range finder would have been useless.

Summarizing, we process the data as follows: 1) IMU is used to detect plane-to-plane transitions; 2) visual odomotery is applied to measure the displacement of two points when a transition occurs; 3) 1D SLAM is performed to extract the number of planes and to cluster data in sub-sets each belonging to a single plane; 4) 2D SLAM is applied for each sub-set of data belonging to a single plane; 5) 2D maps are aligned using visual odometry information computed before. These steps are described in details in the rest of this section.

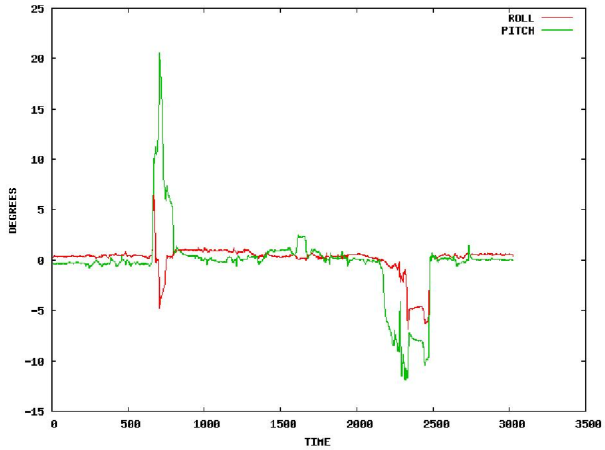

Fig. 1. The *ROLL* and *PITCH* data coming from the IMU sensor. The robot navigates in a planar environment except a single small step (5cm) that the robot has to climb down and then up. In correspondence with these events it is evident a variation in the data coming from the IMU sensor.

A. PLANE TRANSITION DETECTION

To detect a possible change in the plane we analyze data from the IMU sensor. The IMU is collocated on the mobile robot, and we use it to retrieve the ρ,σ Euler angles, respectively the *roll* and *pitch* angles of the robot (in fact, other information would be available, but they are not of interest in this application). An example of the ρ and σ values provided by the IMU is reported in Figure 1. The data refer to a path in which the robot has first to climb down and then climb up a small step (less than 5 centimeters). Even if the depth of the step is small, the data show how clearly the sensor can be used to detect such a change. In fact, it is enough to apply a threshold to the pitch angle σ (the pitch is enough because our robot cannot move sideways) to detect a change. It must be noted that this procedure is affected by false positives. Indeed, when a robot need to overcome an obstacle on its path, there might be a significant change of the σ value. This will be taken into account in the 1D SLAM process by merging planes at the same height (see below).

B. VISUAL ODOMETRY

In order to determine the movement of the robot while it is going through a transition phase, we use a visual odometry technique that processes the input images from the stereo camera mounted on the robot. While the robot is in a transition phase, the data coming from the scanner are not taken into account (thus also preventing the 2D mapping algorithm to take into account the non valid laser data when the robot is climbing over an obstacle). Instead the data coming from the camera are used to determine the movement of the robot.
The visual odometry technique that we have used is based on a standard approach [11]. This process is implemented by using a feature detector to identify feature matches between two consecutive stereo images. Then triplets of features are used to determine the 3D displacement of the camera (and thus of the robot) with a RANSAC approach.

Feature detection is based on the well known KLT feature tracker [12]. For our visual odometry implementation, we consider 200 features tracked over consecutive frames.

To compute the rigid displacement between two consecutive frames f_t and f_{t+1}, in a noise-less case, it would be enough to have three feature associations over the two frames. Having to drop the noise-less assumption, one might consider using a least square method [13], possibly with more than three associations. Nevertheless, the presence of outliers would still represent a major problem that should be taken into account. In order to do this, a RANSAC algorithm [14] is first used to remove outliers. In particular, a set of candidate transformations $\mathbf{T}_i$ from f_t to f_{t+1} are calculated by randomly sampling triplets of feature associations over the two frames. Each transformation $\mathbf{T}_i$ is evaluated by calculating the residual distance

$$d(\mathbf{T}_i) = \sum_{\langle\alpha,\beta\rangle} (\mathbf{T}_i\alpha - \beta)^2$$

where α is a generic feature of frame f_t and β is the associated feature in frame f_{t+1}. The triplets with smallest residual distance are chosen and optimized together to yield the final estimated rigid transformation between the two frames.

Visual odometry process explained above is iterated for a small number of frames (10 to 20 depending on the situation) that are selected in such a way they cover the passage from one plane to another. More specifically, by analyzing IMU data, we can select two time steps: t_S is the starting time of the change of level, i.e., the robot at time t_S is on the first plane, t_E is the ending time of this process, i.e., the robot at time t_E is on the second plane. Then, we consider a set of intermediate frames within this time interval.

It is important to observe here that using visual odometry for a short time allows for ignoring the incremental error that is generated with this method. Moreover, we can further reduce such an error, by using a bundle-adjustment approach, like the one proposed in [15], that considers not only consecutive frames but also frames that are distant in time to improve the quality of the solution.

C. 1D SLAM

The multiplanar mapping could be handled as a series of 2D planar mappings if one could separate the data coming from each of the planes and could know the relative position of one floor level with respect to the others. The problem of calculating this displacement can be termed 1D SLAM, since the unknown value is principally the vertical position z_t of the robot (and, as a consequence, of the different planes). We can model the probelm as

$$z_{t'} = z_t + \Delta z_{[t:t']} \quad (1)$$

where $\Delta z_{[t:t']}$ is the displacement between z_t and $z_{t'}$ calculated from the observations. The problem then becomes that of evaluating $\Delta z_{[t:t']}$. Exploiting again the assumption, it is easy to realize that for most of the times the value of $\Delta z_{[t:t']}$ for two frames close in time f_t and $f_{t'}$ will be zero, since most of the time the robot will be navigating a planar environment. Therefore, it is sufficient to evaluate $\Delta z_{[t:t']}$ while a transition between two planes is occurring. Transitions are detected by using IMU data and measured through visual odometry, as explained before. Therefore $\Delta z_{[t:t']}$ is modeled as

$$\Delta z_{[t:t']} = \begin{cases} 0 & \text{if } |\sigma| < threshold; \\ \Delta z^{\text{VO}}_{[t:t']} & \text{otherwise.} \end{cases} \quad (2)$$

where $\Delta z^{\text{VO}}_{[t:t']}$ is the vertical component of the displacement of the robot position between time t and t' measured with visual odometry and σ is the pitch of the robot measured with the IMU. However, this modeling does not consider the *loop closure* problem, that arises when visiting for a second time a place. In the 1D SLAM problem, this means that the robot can visit the same floor level twice. For example, a robot might explore a room, leave the room by climbing down the stairs, explore another floor level and then, possibly through another entrance, enter again the already visited room by climbing up the stairs or a ramp. Being the visual odometry, and as a consequence the $\Delta z^{\text{VO}}_{[t:t']}$, affected by noise, the z_t will not be the same both times the robot visit the same floor. A procedure to recognize if the floor level has already been visited must be considered. In our case, not having to deal with many different floors, we used a simple nearest neighbor approach. In particular, a new floor g_i is initialized after a change in the level has been detected at time t (and at the beginning of the exploration, of course) and inserted in a set $\mathcal{G}$. The floor is assigned with the measured z_t. Then each floor g_i is checked against every g_j in $\mathcal{G}$ and if the distance is less than a threshold, the two planes are merged and one of them is removed from $\mathcal{G}$. Though the simplicity of the approach, the procedure has been found to successfully merge the same plane when explored twice.

D. 2D SLAM

For each plane g_i in $\mathcal{G}$, a 2D map is computed. In order to do this, a SLAM algorithm [10] is applied on all the laser data collected in each plane. The method simulates a Rao-Blackwellized particle filter by using a hybrid map representation, consisting of small local maps connected together in a graph-like style. The filter is then used to estimate the displacement among these local maps.

Since the different planes have been separated and opportunely merged, there is no need to further develop the 2D SLAM method, that indeed can be applied in its original formulation. The only thing that is necessary to do is to opportunely reinitialize the robot position every time a transition between two planes occurs. This can be done by simply inserting a new link into the graph structure and initialize its distribution by using the variance around the position, estimated with visual odometry. The spreading must be taken into account accordingly to the extent of the stereo measurement noise.

E. MAPS ALIGNMENT

The final process is to align the 2D maps by determining, for each pair of adjacent maps, the displacement of two points in them. Notice that, our assumption is to navigate in a multi-level planar environment, with all parallel planes, thus only 4 parameters are needed to register different 2D maps. Consequently, for each pair of adjacent and consecutive 2D maps, the values Δx,Δy,Δz,$\Delta \theta$ computed by visual odometry are used to align the two maps and a single multi-level map of the environment is thus computed.

V. RESULTS AND DISCUSSION

The system described in the previous sections has been tested on a real scenario integrating data from a 2D laser range finder, an IMU sensor and a stereo camera. The scenario used for the experiments is on three levels: our laboratory, an outside corridor and the street level. The corridor level is about 5 cm below the lab level and the robot had to go down a single step to reach it. Instead the street level is separated from the corridor by a 6 steps stair, about 1 meter high. Due to the mobility limitations of the robotic platform used in the experiments, this level has not been navigated by the robot, but it has only been observed from the corridor level. This limitation could be overcome by just using a more powerful robotic platform. Two different data sets from the same environment have been acquired and processed. The results are similar, so we will report only the ones from the first data set. The size of the explored environment is 18 x 12 meters and the total acquisition time has been 13 minutes. Storage requirements are mainly due to the stereo vision data. In the current implementation, we did not use a compressed format to store stereo data, thus for each stereo image we store on a local disk the left 640x480 color image and the disparity map at 1 frame per second. The total amount of disk space needed to store stereo data has been 1.12 GB, that is about 86.8 MB/min. Data from the LRF and IMU sensors have been acquired at 10 Hz, but disk space used for their storage is very small compared to the stereo data.

As already mentioned, all the map reconstruction processing was performed off-line. The only processing modules that were active on-board were the autonomous exploration module based on a 2D map generated through a simple scan matching method [16] and the stereo correlation algorithm to produce the disparity map [17].

Figure 2 and Figure 3 show some results of our system.

VI. CONCLUSIONS AND FUTURE WORK

In this paper we have proposed a method for reconstructing a piecewise quasi-planar scenario through the use of a laser range finder, a stereo camera and an IMU. First the localization and mapping problem is decomposed in a 1D SLAM method that makes use of the IMU and the stereo camera and a 2D SLAM method that makes use of the laser data. The reconstruction is based on clustering LRF data coming from different planes and on using state-of-the-art 2D

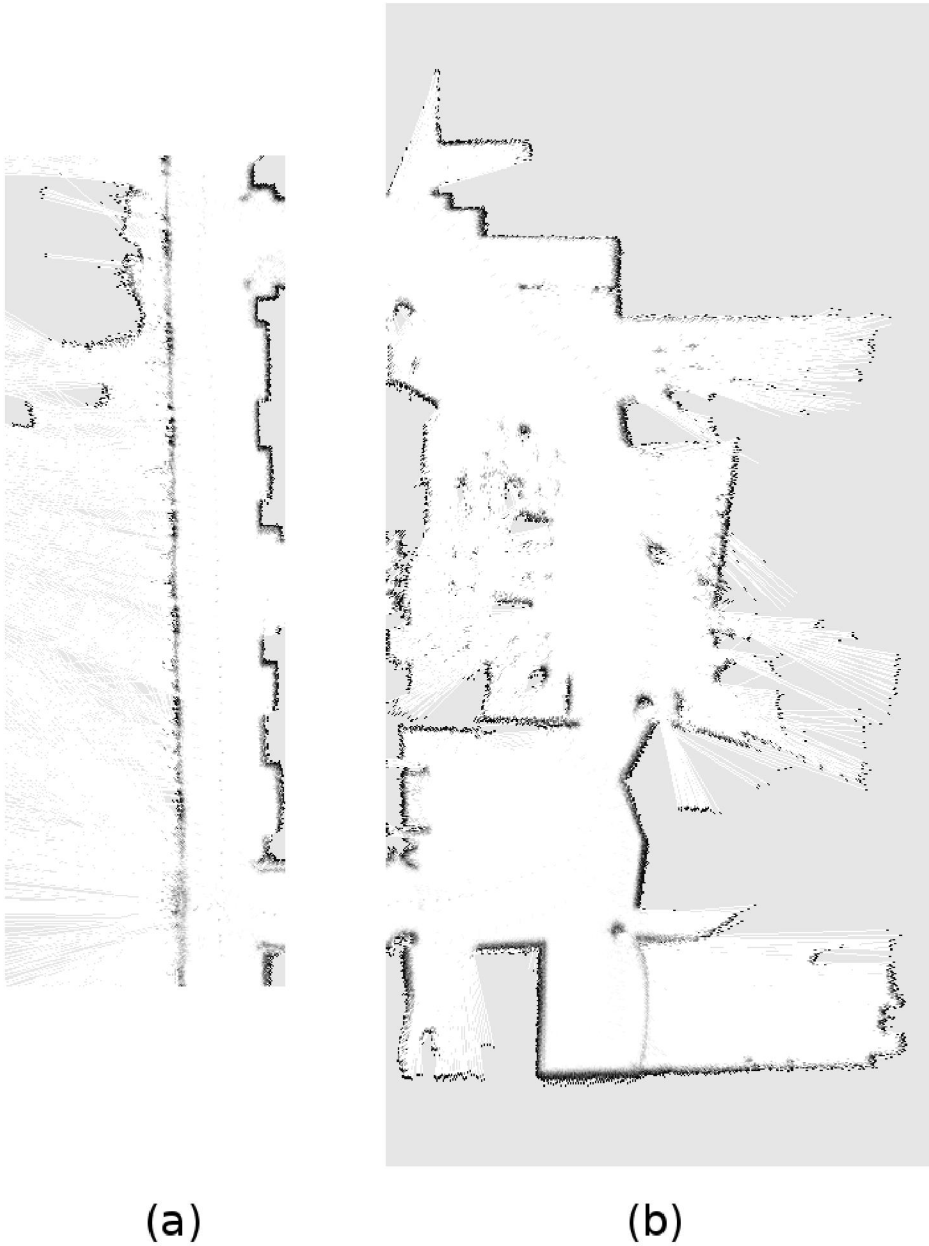

Fig. 2. 2D maps acquired for each different plane.

SLAM methods that are very robust and effective in very large environments.

As future work, we are investigating an extension of the proposed method to be applied on-line during robot exploration.

REFERENCES

[1] J. Diebel, K. Reuterswd, S. Thrun, J. Davis, and G. R., "Simultaneous localization and mapping with active stereo vision," in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2004.

[2] S. Thrun, C. Martin, Y. Liu, D. Hähnel, R. Emery-Montemerlo, D. Chakrabarti, and W. Burgard, "A real-time expectation maximization algorithm for acquiring multi-planar maps of indoor environments with mobile robots," *IEEE Transactions on Robotics and Automation*, vol. 20, no. 3, pp. 433–443, 2004.

[3] A. Früh, C. Zakhor, "An automated method for large-scale, ground-based city model acquisition," *Int. Journal of Computer Vision*, vol. 60, no. 1, pp. 5–24, 2004.

[4] A. Nüchter, K. Lingemann, J. Hertzberg, and H. Surmann, "6D SLAM with approximate data association," in *Proc. of the 12th International Conference on Advanced Robotics (ICAR)*, 2005, pp. 242–249.

[5] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "MonoSLAM: Real-time single camera SLAM," *IEEE Transactions on Pattern Analysis and Machine Iintelligence*, vol. 29, no. 6, 2007.

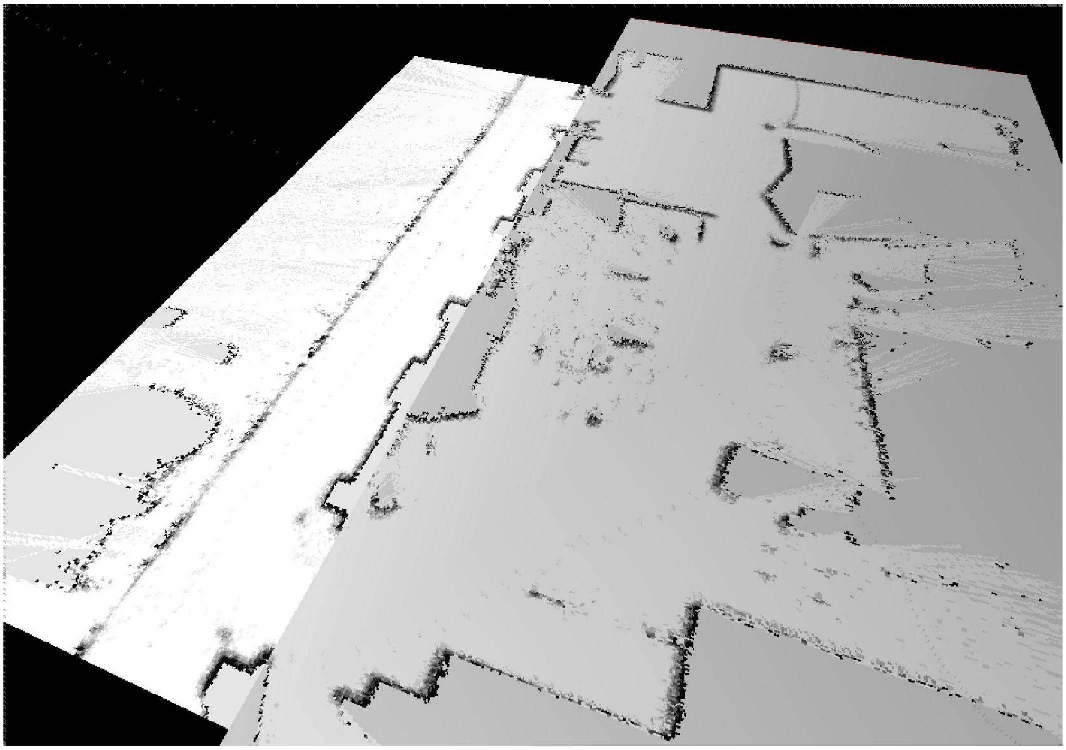

Fig. 3. 3D view of the 2D maps acquired for each different plane.

[6] K. Konolige, M. Agrawal, R. C. Bolles, C. Cowan, M. Fischler, and B. Gerkey, "Outdoor mapping and navigation using stereo vision," in *Proc. of Int. Symposium on Experimental Robotics (ISER)*, 2006.

[7] T. R., P. Pfaff, and W. Burgard, "Multi-level surface maps for outdoor terrain mapping and loop closing," in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, Beijing, China, 2006.

[8] P. Newman, D. Cole, and K. Ho, "Outdoor slam using visual appearance and laser ranging," in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, Orlando, FL, USA, 2006, pp. 1180–1187.

[9] D. Calisi, A. Farinelli, L. Iocchi, and D. Nardi, "Autonomous navigation and exploration in a rescue environment," in *Proceedings of the 2nd European Conference on Mobile Robotics (ECMR)*, Edizioni Simple s.r.l., Macerata, Italy, September 2005, pp. 110–115, iSBN: 88-89177-187.

[10] G. Grisetti, G. D. Tipaldi, C. Stachniss, W. Burgard, and D. Nardi, "Speeding up rao blackwellized slam," in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, Orlando, FL, USA, 2006, pp. 442–447.

[11] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[12] J. Shi and C. Tomasi, "Good features to track," in *IEEE Conference on Computer Vision and Pattern Recognition*, 1994, pp. 593–600.

[13] A. Lorusso, D. W. Eggert, and R. B. Fisher, "A comparison of four algorithms for estimating 3-d rigid transformations," in *BMVC '95: Proceedings of the 1995 British conference on Machine vision (Vol. 1)*. Surrey, UK, UK: BMVA Press, 1995, pp. 237–246.

[14] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Comm. of the ACM*, vol. 24, pp. 381–395, 1981.

[15] K. Konolige and M. Agrawal, "Frame-frame matching for realtime consistent visual mapping," in *Proc. of International Workshop on Robot Vision*, 2007.

[16] G. Grisetti, "Scaling rao-blackwllized simultaneous localization and mapping to large envronments," Ph.D. dissertation, University of Rome 'La Sapienza', Dipartimento Di Informatica e Sistemistica, 2006.

[17] K. Konolige, "Small vision systems: Hardware and implementation," in *Proc. of 8th International Symposium on Robotics Research*, 1997.