<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper presented at *15th Annual Conference, TAROS (Towards Autonomous Robotic Systems) 2014, Birmingham, United Kingdom, September 1-3, 2014.*

Citation for the original published paper:

Gholami Shahbandi, S., Åstrand, B. (2014)

Modeling of a Large Structured Environment: With a Repetitive Canonical Geometric-Semantic Model.

In: Michael Mistry, Aleš Leonardis, Mark Witkowski & Chris Melhuish (ed.), *Advances in Autonomous Robotics Systems: 15th Annual Conference, TAROS 2014, Birmingham, UK, September 1-3, 2014. Proceedings* (pp. 1-12). Heidelberg: Springer

Lecture Notes in Computer Science

http://dx.doi.org/10.1007/978-3-319-10401-0_1

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:hh:diva-26316>

Modeling of a Large Structured Environment

With a Repetitive Canonical Geometric-Semantic Model ^{*}

Saeed Gholami Shahbandi and Björn Åstrand

Center for Applied Intelligent Systems Research (CAISR),
Intelligent Systems Lab, Halmstad University, Sweden

Abstract. AIMS project attempts to link the logistic requirements of an intelligent warehouse and state of the art core technologies of automation, by providing an awareness of the environment to the autonomous systems and vice versa. In this work we investigate a solution for modeling the infrastructure of a structured environment such as warehouses, by the means of a vision sensor. The model is based on the expected pattern of the infrastructure, generated from and matched to the map. Generation of the model is based on a set of tools such as closed-form Hough transform, DBSCAN clustering algorithm, Fourier transform and optimization techniques. The performance evaluation of the proposed method is accompanied with a real world experiment.

1 Introduction

Following the advances of the state of the art in autonomous vehicles, and growing research on the design and development of innovative solutions, *intelligent warehouses* emerge leveraging insights from several specialist domains. The Automatic Inventory and Mapping of Stock (AIMS) project targets the traditional warehouses where not necessarily infrastructures are designed or installed for automation. AIMS project intends to develop a process, through which an awareness of the surrounding environment is embedded in a “live” semantic map, for effective management of logistics and inventory (see fig. 1). Achieving this objective requires different sensors contributing to this semantic map in multiple layers. Forklift trucks enriched with such an awareness enable safe and efficient operations while sharing the workspace with humans. In such a shared workspace, compatibility between vehicles’ knowledge, humans’ cognition and Warehouse Management Systems (WMS) is important.

This paper focuses on mapping and modeling the infrastructure of the warehouse, as a foundation for addressing the location of both vehicles and storage of the warehouse. We present a method to extract structural pattern of the map that serves as the foundation of a multilayer geometric-semantic map to be used for logistic planning, Auto Guided Vehicle (AGV) path planning and WMS interaction. Semantic labels are subject to the context, in order to be functional

^{*} This work as a part of AIMS project, is supported by the Swedish Knowledge Foundation and industry partners Kollmorgen, Optronix, and Toyota Material Handling Europe.

Fig.1: The objective in AIMS project is to provide an awareness framework. The framework interacts with Warehouse Management System (WMS).

they shall be defined based on the environment and application. Accordingly we integrate multiple layers in this map, from geometric layout design of the environment to semantic annotation of the regions and infrastructures as a model. For this purpose a map of pillars of the pallet rack cells (see fig. 1) is generated, from a fish-eye camera based on a bearing-only mapping method. This map is used for inference and generation of a layout map of the warehouse, which integrates the desired conceptual meaning. Contribution of this work is a semantic-geometric model, based on an object map entailing infrastructure elements as landmarks. This is followed by an easement in inferences through a method for extracting and matching the model from and to the aforementioned map. Reliability and performance of the proposed model and accompanying method is demonstrated on a map acquired from a real world warehouse.

1.1 Related Works

In a time where robots are embedded in our daily life, robot's awareness of their surrounding is a crucial competence and semantic mapping is a particularly important aspect of it. That is because, while a geometrically consistent map is sufficient for navigation, it is not enough for task planning and reasoning. Many researchers have been contributing to this particular aspect, from different points of view.

Some tried to model the environment through the topology of open space in geometric map, like [6], where they employed a series of kernels for semantic labeling of regions. Some others like [5], [14] and [9] proposed spatial maps, enhanced conceptually by object recognition in regions of the map. [5] proposed a composition of two hierarchical maps, semantic and geometric anchored together. In [14] a framework of a multilayer conceptual map is developed, representing the spatial and functional properties of the environment. And [9] introduced a comprehensive framework of spatial knowledge representation for large scale semantic mapping.

While mentioned researchers aimed to derive semantic concept from the functionality of the objects into the map, some others such as [8], [12] and [7], in-

roduced properties of the regions as semantic label. [8] annotates an occupancy map with the properties of the regions, either “building” or “nature”, through data from range scanner and vision. In [12] properties of the environment such as terrain map and activity map, are embedded into a metric occupancy map. Concerning the global localization, [7] employed hybrid geometric-object map.

Mentioned works are proposed for cases where the global structure of the environment is not a concern, and semantic information is extracted locally. The conceptual meaning is the property or functionality of the content of those regions, and does not link to the structural model of the environment. Researchers have also taken into account the environment’s structure. [13] attempted mastering the SLAM problem by a geometrical object-oriented map representation, through modeling the boundaries of obstacles with polygons, employing a Discrete Curve Evolution (DCE) technique. An interesting recent work [10], developed a method for detection, evaluation, incorporation and removal of structure constraint from latent structure of the environment into a graph-based SLAM. Two last examples take into account the structure of the environment, for improving the solution to SLAM problem and providing a more consistent map. However there is no conceptual meaning accompanying the extracted structure.

Spatial semantic of open space from occupancy map is an interesting aspect and we investigate it in another work. It does provide useful knowledge of the open space in a warehouse, such as connectivity, corridors, or crossing of the corridors. But it does not provide semantic labels for infrastructure, such as the entrance of a pallet rack cell (see fig. 2). Such an information is very useful when the articles are localized in the layout, for logistic and AGVs’ task planning. The other approaches to semantic mapping, where the semantic labels are derived from objects in the region is not very beneficial to our work either. That is because the smallest entities of regions are pallet rack cells with same semantics. Stored articles in those pallet rack cells and their identities do not carry any conceptual meaning for their region. The objects that we are interested in are the infrastructure of the warehouse, such as pillars of the pallet rack cells which represent the structure of the environment (see fig. 2). Therefore a more suitable approach for us would be to create a map of the environment, using the infrastructure as landmarks, and then extracting those patterns that are meaningful for us.

1.2 Our Approach

This paper presents a canonical geometric model and describes how to match it into the latent structure of a map. Such a model enables the further processing of geometric-topological modeling of the environment, for the purpose of semantic annotation of structures and geometric layout extraction. Semantic concept is encoded into the model through the choice of landmarks in the map as shown in fig. 2. It is assumed that the environment is highly structured and a pattern is frequent enough, so that it is possible to effectively represent the whole structure by a set of this canonical model with different parameters. By choosing pillars as landmarks, opening of a pallet rack cell is implied by two neighboring landmarks, while the sequence of landmarks creates a layout map by the “boundaries” of the corridors (a similar concept of representation as in

Fig. 2: Design of the model derived from the map’s context, encoding semantic-geometric relation of landmarks.

[13]). This explains how the model in fig. 2 embeds a sufficient level of semantic and geometric knowledge while it is expressed only by a set of points uniformly distributed on a straight line. Different models are uniquely identified by the means of 5 parameters n, d, x, y, θ . The model and matching method is designed to represent independent lines of landmarks, therefore parameters of different models are independent. However the repetition of the model in the map of a warehouse, makes it possible to pose a global constraint on its parameters in order to achieve a set of models which are globally consistent.

In next section (2), the model and the matching method are explained in details. The process is demonstrated on a synthesized data in order to sketch the generality of the method where the models in one map are completely independent. Section 3 contains the method we adapt for mapping the environment from a fish-eye camera. Resulting map is modeled by the proposed method, while introducing global constraints for a real world map.

2 Model Generation and Fitting

Let’s assume a $\mathcal{MAP}$ consisting of landmarks is given, where each landmark is described by its pose ($\mu = x, y$) and corresponding uncertainty modeled with a 2D normal distribution (with covariance Σ) as defined in equation 1.

$$\mathcal{MAP} = \{lm_i \mid lm_i := \mathcal{N}_i(X)\} \quad , \quad \mathcal{N}_i(X) = \frac{\exp(-\frac{1}{2}(X-\mu_i)^T \Sigma_i^{-1}(X-\mu_i))}{\sqrt{(2\pi)^2 |\Sigma_i|}} \quad (1)$$

The expected pattern in this map was described and motivated in the introduction (section 1). This model as illustrated in fig. 2, represents a set of landmarks (pillars) aligned on one side of a corridor. We call these sets of landmarks $\mathcal{L}$, and we try to fit one model per set.

$$\mathcal{L}_i = \{\{lm_{ij}\}, \theta_i \mid lm_{ij} \in \mathcal{MAP}\} \quad (2)$$

In the definition of $\mathcal{L}_i$ (equation 2) lm_{ij} are landmarks aligned on the side a corridor, and θ_i is the angle of alignment. Model M shown in fig. 2c and expressed in equation 3 is designed to represent sets of landmarks $\mathcal{L}$, and each model is uniquely identified by 5 parameters (n, d, x, y, θ) , where n is the number

of landmarks in the corresponding $\mathcal{L}$, d is the distance between two consecutive landmarks, (x, y) are the coordination of the 1st landmark, and θ_i is the angle of alignment of the set.

$$M_i(n_i, d_i, x_i, y_i, \theta_i) := \{p_{ij}\} \quad , p_{ij} = \begin{bmatrix} x_i + n_j d_i \cos \theta_i \\ y_i + n_j d_i \sin \theta_i \end{bmatrix}, 0 \leq n_j < n_i \quad (3)$$

First step in generating the models M is to segment the $\mathcal{MAP}$ into the sets of landmarks $\mathcal{L}_i$. For this purpose we have developed a closed-form of Hough transform in combination with a clustering algorithm. This technique not only locates the desired $\mathcal{L}$ by (θ, ρ) , but also directly clusters the landmarks into different $\mathcal{L}$. After the segmentation of landmarks into sets, each set is projected into an axis passing through that set (see fig. 4b). This operation will map the 2D normal distributions of the landmarks into a 1D signal. An analysis of the resulting 1D signal in frequency domain will result in an estimation of the n and d of the model. Given the θ, n and d of the model by closed-form Hough transform and frequency analysis, x and y remain for extraction. An optimization would serve this purpose, where the first point of the set $\mathcal{L}$ serves as the initial guess of the optimization process.

2.1 Segmentation by Closed-Form Hough Transform

Transformation of a point from Cartesian space (x, y) into Hough space [3] (θ, ρ) is performed by equation 4.

$$\rho = x \cos \theta + y \sin \theta \quad (4)$$

Conventional form of Hough transform is applied to discrete images. Hough space is also a discretized image where the value of each pixel (θ, ρ) is given by the summation of the value of all pixels (x, y) that satisfy the equation 4. The peaks in the Hough space represent lines in original image where points are aligned. But we are interested in more than that. We would like to know which particular points in Cartesian space contributed to an specific peak (θ, ρ) in Hough space. Therefore we introduce a closed-form solution of Hough transform to address that issue. This is realized by representing each point (x_i, y_i) , with a corresponding sinusoid as expressed in equation 4. Next step is to intersect all resulting sinusoids and store the intersections with $(\theta_{ij}, \rho_{ij}, i, j)$. Where (θ_{ij}, ρ_{ij}) represent the location of the intersection in Hough space and i, j are the indices of intersecting sinusoids. Outcome of this step is a set of intersection points in Hough space as illustrated in fig. 3b.

Advantages of closed-form approach are, first, clustering the intersection points in Hough space directly corresponds to clustering aligned points in Cartesian space. Secondly, it prevents us from discretization of a continues Cartesian space.

From the Hough space, clustering the intersection points is straightforward. We employed the Density-based spatial clustering of applications with noise (DB-SCAN) [4] algorithm. This algorithm requires a value for the minimum number

Fig. 3: Closed-form Hough transform maps the points from Cartesian space into sinusoids in Hough space. A clustering is performed over intersections in Hough space (red circles) to cluster the aligned points in Cartesian space.

of samples per cluster, which is derived from the number of expected points in each alignment $\mathcal{L}$. Assuming a set of n point aligned in Cartesian space ($\mathcal{L}$), consequently there will be the same number of sinusoids in Hough space intersecting $\frac{n(n-1)}{2}$ times at the same location (θ, ρ) . As a perfect alignment in noisy data is unlikely, minimum number of samples are set to half of that ($\frac{n(n-1)}{4}$) to guarantee a successful clustering.

The result of clustering in Hough space is mapped to Cartesian space, generating sets of points $\mathcal{L}_i$ as desired. Angle of alignment $\theta_i \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, is therefore the first estimated parameter of the model.

2.2 Frequency and Length via Fourier Transform

In order to estimate two parameters of the model n, d , each set of landmarks $\mathcal{L}_i$ is projected to an axis passing through the center of mass, projecting the set of 2D normal distribution functions into a 1D signal. This projected signal is illustrated in fig. 4b and defined in equation 5.

$$\begin{aligned} signal_i &= \sum_{lm_j \in \mathcal{L}_i} f_{lm_j}(\mu_j, \sigma_j) \\ f_{lm_j}(\mu_j, \sigma_j) &= \frac{1}{\sigma_j \sqrt{2\pi}} e^{-\frac{(x-\mu_j)^2}{2\sigma_j^2}}, \mu_j = \frac{\vec{v}_1 \cdot \vec{v}_2}{\|\vec{v}_1\|}, \sigma_j = \sigma_x \\ \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix} &= R(\theta_j) \Sigma_j R^T(\theta_j) \end{aligned} \quad (5)$$

In equation 5 vectors $\vec{v}_1, \vec{v}_2$ are those illustrated in fig. 4b. $\vec{v}_1$ is the projection axis itself and $\vec{v}_2$ is a vector from starting point of $\vec{v}_1$ to the point subjected to projection.

Assuming a uniform distribution of points in each set implies that, Fourier transform of the projection signal ($\mathcal{F}(f)$) has a dominant frequency. This frequency relates to the values n, d as expressed in equations 6.

$$\begin{aligned} n &= f_0, d = \frac{\|\vec{v}_1\|}{f_0} \\ f_0 &= \arg \max \mathcal{F}(f) := \{f_0 \mid \forall f : \mathcal{F}(f) \leq \mathcal{F}(f_0)\} \end{aligned} \quad (6)$$

2.3 Optimization and Model Fitting

Last parameters to uniquely define the model are (x, y) . For that purpose, (x, y) are set to the pose (μ) of the first landmark in $\mathcal{L}$. Considering that the angle of alignment θ belongs to $[-\frac{\pi}{2}, \frac{\pi}{2}]$, first element of each $\mathcal{L}$ is the most left item, or the lowest in case of vertical lines. Then through an optimization process (x, y) are tuned. It should be noted that the first landmark in $\mathcal{L}$ is an initial guess for the optimization. This is based on the assumption that $\mathcal{L}$ does not consist of very far off outliers. Such outliers may bring the optimization into a local minima, causing a shift in model's position. Objective function of the optimization is given in equation 7. This function is a summation of all normal distribution functions of landmarks of the line ($\mathcal{N}_{ik} \in \mathcal{L}_i$) operating on all the points of the model ($p_{il} \in M_i$).

$$f_i(X) = \frac{1}{\sum_l \sum_{k \in \mathcal{L}_i} \mathcal{N}_{ik}(p_{il})} \quad (7)$$

Since in the map of a real environment most of the models share 3 parameters (n, d, θ) , for demonstrating the performance and generality of the method, it was applied to a synthesized data instead of a real map. Performance of the proposed method is evaluated over a map from real world in section 3. Unlike real environment the synthesized data in Figure 4a contains 3 $\mathcal{L}$ with all different parameters.

Fig. 4: Modeling 3 segmented line sets ($\mathcal{L}$) from 4a to 4c. 4b shows 2D to 1D projection of a single $\mathcal{L}$ according to equation 5 for frequency analysis

Result of clustering is encoded in colors. While in a real world map a crossing between $\mathcal{L}$ is not expected, in this synthesized data the lines are crossing to demonstrate the performance of the closed-form Hough transform. In addition to a uniform noise added to the position of each landmark, 10% of landmarks are removed to evaluate the result of Fourier transform in estimating n and d . As it is observed in fig. 4c, the model would fill in the position of missing landmarks. Finally the values of the parameters are successfully computed by the method, in fig. 4 M_1 (20, 2.12, -0.1, 0.0, 0.78) colored red, M_2 (20, 1.50, 0.0, 23.0, 0.00) colored green and M_3 (15, 2.80, 0.0, 28.5, -0.78) colored cyan. It should be men-

tioned that in case of a real map not all minor assumptions are met, so global constraints are introduced over all parameters. However none of the global constraints are preset, but all are extracted from the map. This aspect is explored further in section 3.1.

3 Experimental Results and Discussion

This section describes the procedure of creating and modeling a map of a real warehouse. For this purpose we use a “Panasonic” 185° fish-eye lens mounted on “Prosilica GC2450” camera, installed on a AGV forklift truck in a warehouse. As the truck is provided with a localization system based on lasers and reflectors, we adopt an Extended Kalman Filter based bearing-only technique for mapping. Then the model developed in this work is employed to represent the map. To this end, a set of global estimations are extracted from the map and posed over the model’s parameters as global constraints.

Pillar Detection Considering the common color coding of the pillars in the warehouses, pillar detection starts with segmentation through color indexing [11], followed by calculation of the oriented gradient. The camera is pointing downward and all the pillars are parallel to the camera axis. Consequently pillars are pointing to the vanishing point of the camera in the images, hence gradient vectors of pillars’ edges are perpendicular to lines passing through the vanishing point. Any other gradient vectors are considered non-relevant and filtered out. The gradient image is sampled over multiple concentric circles as illustrated in fig. 5b. Results of all sampling are accumulated in a signal as in fig. 5c, capturing the pattern of a pillar’s appearance by two opposite peaks representing the edges of the pillar. Detecting the position of such a pattern returns the bearing to pillars. A continuous wavelet transform (CWT) based peak detection technique [2] is adopted for detection of pillars’ pattern in the gradient signal. The method is based on matching a pattern encoded in a wavelet, by the help of CWT. The pattern representing two sides of the pillar in the gradient signal (see fig. 5c), could be modeled with a wavelet based on the 1st order derivative of a Gaussian function as illustrated in fig. 5d.

Fig. 5: Detection of pillars. 5a) original image and the result of pillar detection, 5b) circular sampling of gradient image. 5c) accumulation of circular sampling as a signal, illustrating the occurrence of pillars. A wavelet in 5d, resembling the pattern of pillars in gradient signal, is used for pillar detection by CWT technique.

Bearing Only Mapping Mapping is performed in an Extended Kalman Filter (EKF) fashion proposed in [1] for bearing-only SLAM. However, as the pose of the truck is provided through a laser-reflector based system, there is no need to localize the truck through EKF framework. Nevertheless, we include the state of the vehicle in the Kalman filter’s state to ease the adaptation and system description. The truck’s state is neither predicted nor updated in corresponding phases of the EKF, instead it comes directly from the localization system. This ensures that Kalman gain computation and prediction step are based on the “true” pose of the truck, and leaves the possibility to extend this implementation into SLAM if required.

A map generated through this framework, from the data logged in a real world warehouse is sketched in fig. 6. Colors in this map code the result of clustering of the line sets ($\mathcal{L}$), and blacks are outliers.

Fig. 6: Mapping 77 lines of pillars of a warehouse of size $250^m \times 72^m$.

3.1 Global Constraints on the Models

Recalling the assumption of a structured environment, the set of models representing the boundaries of corridors are parallel, and have similar parameters n, d, θ . Even the starting points of the models are aligned on a straight line (like in fig. 6). We exploit this information to pose a set of global constraints, which results in a globally more consistent set of models of the environment.

Dominant Orientation refers to the fact that corridors in an environment have the same direction. This suggest that we could pose a global constraint on θ . This has a crucial importance for the clustering step as well, as the number of intersection points in Hough space (see fig. 3b) increases quadratically by the number of landmarks ($\frac{n(n-1)}{2}$). This means for a map consisting of more than 2000 landmarks like fig. 6, there will be more than 2 million intersection points in Hough space. Such a number of points can not be easily clustered. However, considering the assumption of dominant orientation, it is possible to reject a huge number of intersection points which are not within an acceptable range (5°) from dominant orientation. This filtering process is handled by discretization of the Hough space

and finding the dominant orientations, where most of the peaks occur. After the filtering process, the clustering proceeds as explained earlier in closed-form.

Dominant Frequency Similarly a constraint is posed over parameters n and d . This is also very helpful since in a real map, sometimes landmarks are not visible and hence not included in the map. If the percentage of missing landmarks become too big, it may yield a wrong estimation of the frequency. In order to estimate global constraint of these two parameters, first n and d are computed for all lines ($\mathcal{L}$) without constraint. Then the mode in the histogram of these parameters as shown in fig. 7 provides a global estimation of the parameters. After acquiring the global values from the map, models are generated again, this time with the global constraints.

Initialization Line Assumption of a structured environment implies that the starting points of all $\mathcal{L}$ are located on the same line. This line is then used for initial guess in the optimization process. The initialization line is calculated by a linear regression among first landmarks in all $\mathcal{L}$, shown in fig. 7c. While each $\mathcal{L}$ comes from Hough space with its line equation, the exact position of the optimization’s initial guess for each $\mathcal{L}$ is given by the intersection point of mentioned line and initialization line.

Fig.7: Global constraints; Histograms of (n,d) , mode in each shows the dominance. Initialization line is a linear regression of first landmarks in all $\mathcal{L}$ (red points).

3.2 Discussion

Engaging all the global constraints, result of modeling the map given in fig. 6 is illustrated in fig. 8a. The final result of modeling is useful for geometric mapping of the boundaries of corridors, as well as semantic annotation of pallet rack cell on the side of corridors. A crop of the model augmented with truck’s pose and reflectors is presented in fig. 8b. The reflectors are those pre-installed in the environment used for AGVs’ localization. We use the position of reflectors as ground truth for accuracy estimation. A box plot in fig. 8c demonstrates the distances

between 1143 reflectors and nearest point in the model. Taking into account the width of pillar's and reflector's mount, the distances are representing the errors of the model with an offset of about 1 decimeter.

Fig. 8: Modeling a warehouse (see fig. 6) with geometric semantic models (all axes' unit in meter). 8b) shows a crop of model augmented with the truck's pose (black arrow) and reflectors (white triangles). 8c) shows distance between models and closest reflectors, with an offset about 1^{dm} .

It must be remembered, although the global constraints improve the global consistency of the models, yet it is only applicable if the infrastructures share similar geometry, that is to say the resulting models have the same parameters. In case of an environment that consists of multiple *regions* with different characteristics, those regions must be subjected to regional constraints.

4 Conclusion

Toward developing an awareness framework for effective management of logistics and inventory in a warehouse environment, a rich geometric-semantic map is set as a goal. In this work we address this objective on the level of infrastructure modeling. We suggest a choice of landmark for bearing-only mapping of a warehouse environment, leading to a map which integrates the infrastructure of the warehouse. This paper presents a canonical geometric-semantic model, along with a method for generating and matching these models into the map. The model is a set of points aligned on a straight line uniquely identified by 5 parameters. Geometrically it represents the boundaries of corridors, and semantically represents pallet rack cells. A series of tools such as closed-form Hough transform, DBSCAN clustering algorithm, Fourier transform and optimization techniques are employed to estimate the parameters of each model and match them with the map. The performance of the method is demonstrated over both synthesized data and real world map.

Proposed model and extracting method have interesting characteristics, such as modularity, generality, and representing both geometric and semantic knowledge. Most of the steps and parameters of the model are independent, therefore it is adjustable according to different scenarios. One example of such adjustment was given as global constraint in this work, where all of the models' parameters adopt to the global characteristic of the map for a better consistency. However

the method has its own limits as discussed in 3.2, such as sensitivity to initial guess of the optimization, and consideration of global constraints for consistency.

We plan to develop this work further by fusing other sensory data into this map, such as introducing occupancy notation from range scanners. And also by developing a more dynamic model, allowing harmonics of the dominant frequency to participate, and handling those cases where landmarks are not uniformly distributed. Segmentation of map into regions based on dominant orientations would be helpful, if the environment consists of multiple orientations. Indeed we investigated this aspect in another work based on occupancy maps, and will introduce it to this method as soon as the occupancy notation is fused into this map.

References

1. Bailey, T.: Constrained initialisation for bearing-only slam. In: Robotics and Automation, 2003. Proceedings. ICRA'03. IEEE International Conference on. vol. 2, pp. 1966–1971. IEEE (2003)
2. Du, P., Kibbe, W.A., Lin, S.M.: Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics* 22(17), 2059–2065 (2006)
3. Duda, R.O., Hart, P.E.: Use of the hough transformation to detect lines and curves in pictures. *Communications of the ACM* 15(1), 11–15 (1972)
4. Ester, M., Peter Kriegel, H., S, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. pp. 226–231. AAAI Press (1996)
5. Galindo, C., Saffiotti, A., Coradeschi, S., Buschka, P., Fernandez-Madrigal, J.A., González, J.: Multi-hierarchical semantic maps for mobile robotics. In: Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on. pp. 2278–2283. IEEE (2005)
6. Liu, Z., von Wichert, G.: Extracting semantic indoor maps from occupancy grids. *Robotics and Autonomous Systems* (2013)
7. Park, S., Kim, S., Park, M., Park, S.K.: Vision-based global localization for mobile robots with hybrid maps of objects and spatial layouts. *Information Sciences* 179(24), 4174–4198 (2009)
8. Persson, M., Duckett, T., Valgren, C., Lilienthal, A.: Probabilistic semantic mapping with a virtual sensor for building/nature detection. In: Computational Intelligence in Robotics and Automation, 2007. CIRA 2007. International Symposium on. pp. 236–242. IEEE (2007)
9. Pronobis, A., Jensfelt, P.: Large-scale semantic mapping and reasoning with heterogeneous modalities. In: Robotics and Automation (ICRA), 2012 IEEE International Conference on. pp. 3515–3522. IEEE (2012)
10. de la Puente, P., Rodriguez-Losada, D.: Feature based graph-slam in structured environments. *Autonomous Robots* pp. 1–18 (2014)
11. Swain, M.J., Ballard, D.H.: Color indexing. *International journal of computer vision* 7(1), 11–32 (1991)
12. Wolf, D.F., Sukhatme, G.: Semantic mapping using mobile robots. *Robotics, IEEE Transactions on* 24(2), 245–258 (2008)
13. Wolter, D., Latecki, L.J., Lakämper, R., Sun, X.: Shape-based robot mapping. In: KI 2004: Advances in Artificial Intelligence, pp. 439–452. Springer (2004)
14. Zender, H., Martínez Mozos, O., Jensfelt, P., Kruijff, G.J., Burgard, W.: Conceptual spatial representations for indoor mobile robots. *Robotics and Autonomous Systems* 56(6), 493–502 (2008)