Preprint

This is the submitted version of a paper presented at *13th International Conference on Control, Automation, Robotics and Vision, ICARCV 2014, Marina Bay Sands, Singapore, December 10-12, 2014.*

Citation for the original published paper:

Gholami Shahbandi, S., Åstrand, B., Philippsen, R. (2014)
http://dx.doi.org/10.1109/ICARCV.2014.7064584

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-26597
Sensor Based Adaptive Metric-Topological Cell Decomposition Method for Semantic Annotation of Structured Environments

Saeed Gholami Shahbandi  
saesha@hh.se  
Halmstad University, Sweden  
Box 823, 30118 Halmstad

Björn Åstrand  
bjorn.astrand@hh.se  
Halmstad University, Sweden  
Box 823, 30118 Halmstad

Roland Philippsen  
roland.philippsen@hh.se  
Halmstad University, Sweden  
Box 823, 30118 Halmstad

Abstract—A fundamental ingredient for semantic labeling is a reliable method for determining and representing the relevant spatial features of an environment. We address this challenge for planar metric-topological maps based on occupancy grids. Our method detects arbitrary dominant orientations in the presence of significant clutter, fits corresponding line features with tunable resolution, and extracts topological information by polygonal cell decomposition. Real-world case studies taken from the target application domain (autonomous forklift trucks in warehouses) demonstrate the performance and robustness of our method, while results from a preliminary algorithm to extract corridors, and junctions, demonstrate its expressiveness. Contribution of this work starts with the formulation of metric-topological surveying of environment, and a generic n-direction planar representation accompanied with a general method for extracting it from occupancy map. The implementation also includes some semantic labels specific to warehouse like environments.

I. INTRODUCTION

The state of the art in autonomous robotics has advanced sufficiently that open implementations of many core technologies are now readily available. Consequently, there is growing research on the design and development of innovative solutions that leverage insights from several specialist domains. The Automatic Inventory and Mapping of Stock (AIMS) project lies in this category targeting the traditional warehouses where no infrastructure is designed or installed for automation. Its goal is to develop a system that seamlessly combines inventory management with autonomous forklift trucks in intelligent warehouses (Fig. 1). Information compatible with human operators, management systems, as well as mobile robots is of particular importance here. A rich and “live” map combining metric and semantic information is a crucial ingredient for effective management of logistics and inventory, especially for autonomous fleets working in the same space as humans and human-operated devices.

This paper focuses on a subsystem developed for AIMS, namely a robust method to automatically analyze an occupancy grid map.

It produces a “subdivision” representation, whose elements correspond to a variety of spatial resolutions and orientations that are the most relevant for a given environment and application context. This representation serves the task of surveying a new environment, as input to a first stage of semantic annotation based on spatial features for further analysis. The resulting representation is expected to later feed into CAD softwares for setting up and operating Auto Guided Vehicles (AGV) in warehouses, as well as interact with Warehouse Management Systems (WMS).

Fig. 1: The AIMS project establishes a set of tools for setting up and operating AGVs such as forklift trucks. It includes a range of aspects, from surveying of warehouses to on-line inventory tracking.

The environment is assumed to have a given number of predominant directions, but the angles between them are not constrained. We also assume that the spatial scale of the semantic features to be detected in the subsequent step is approximately known. To illustrate these points and demonstrate the suitability of our method, we also describe a preliminary algorithm to detect corridors and junctions. Experiments based on two real-world data sets from different sources show the versatility and robustness of our method.
There is a significant body of work on semantic analysis of the kinds of environments relevant for AIMS. We discuss only the ones most pertinent to this paper. Local approaches base semantic information either on object labeling or the identification of local shapes. For instance, classifiers can be used to detect regions in range data [1]. Image features, and similarly range features, can also be used [2]. To label segments in 3D point clouds, visual appearance, shape, geometry, and geometrical context can be used [3]. Global approaches on the other hand derive semantic information via place labeling. For instance, [4] builds a topological map based on the notion of connectivity and adjacency. [5] explores an occupancy grid map through the Voronoi graph. [6] bases a topological map on the connectivity between acquired images. [7] provides a topological graph of the environment based on the concept of virtual door, and [8] employs a series of kernels based on Markov chain Monte Carlo for semantic labeling.

For our purposes, a hybrid metric-semantic approach is most appropriate. Our formulation relies heavily on topological aspects, while grounding the representations in metric sensor-based information, and leveraging both in the abstraction and labeling algorithms.

**Contributions and Approach**

The common reliance on line extraction [9], [8], [10], [11], [12] indicates the importance of orientation in semantic mapping. However, we cannot readily reuse these results, due to clutter and discontinuities in the physical structures. For instance, the walls of warehouse corridors consist of irregularly stored articles on shelves built with pillars whose horizontal extents are significantly smaller than the goods containers.

Thus, a particular challenge that our method addresses is the discrepancy between the importance of orientation information and the absence of corresponding clean long lines in the range data that can be obtained via laser scanners in warehouses. We achieve this with a combination of Histogram of Oriented Gradients [13] and a Radiogram [14] operating on an occupancy map. The radiogram is similar to a Radon Transform [15], operating on specific angles and associated with a filtering mechanism based on oriented gradients.

The histogram of oriented gradients (Fig. 2) allows to extract a (given number of) dominant directions in the environment. Radiography is then applied to occupied cells in direction of dominant orientations, and the resulting peaks correspond to feature locations in the map. The real-world environments shown in this paper exhibit two perpendicular dominant orientations, but our method is by no means limited to these cases. Employing radiography based technique for feature extraction requires the resolution of the map to be sufficiently high to represent each side of the physical entities with enough observations. The spatial arrangement of features is extracted via polygonal cell decomposition, and captured in a subdivision and an adjacency graph data structure to enable a higher-level topological analysis.

We collect the information on orientations, locations, the subdivision and adjacency graph data structures in our adaptive method. To clarify terminology, adaptivity here is with respect to the geometric structure of the environment while preserving the metric information. This is distinct from resolution adaptivity such as found in quadtrees or octrees [16], [17].

Our novel representation characterizes the geometric structure of the environment in addition to providing an abstract metric representation of the map.

Note that this paper does not fully address subsequent semantic analysis, but focuses on an adaptive generic representation to support computing and storing of such info, and one specific implementation of the method for environments with two dominant directions. Developing a more complete and versatile annotation method and representation is part of ongoing and future research.

**II. Method**

Performing a survey by means of an occupancy grid map demands an understanding of the general structure of the map. We rely on the assumption that there exist physical elements approximately arranged in straight lines according to an arbitrary limited set of dominant directions. Our method consists of the following main steps.

1) Detect dominant directions via histograms of orientated gradients in the occupancy map (Fig. 2, section II-A).
2) Extract line features via radiography along the dominant directions, using wavelet-based peak detection to robustly influence the resulting resolution (Fig. 3a, section II-B).
3) Employ polygonal cell decomposition to compute a planar subdivision and an adjacency graph (Fig. 3b, section II-C).
4) Refine the adjacency abstraction by analyzing occupancy inside faces and along edges (Fig. 5, section II-D).
5) Combine metric and topological information to infer semantic labels using template matching (Fig.8, section II-E).

In this paper, we focus on producing the subdivision, which captures the structure of physical entities in the environment, as well as the adjacency graph, which encodes the topology (connectedness of open space). Inference is sketched for the kind of environment of current relevance to practical advances in the AIMS project, looking at two usage scenarios: aiding in setting up AGV operation, and online path planning.

The method is demonstrated first on a map of the Intel Jones Farms Campus\(^1\) and later applied to data collected in the warehouse of one of our industrial partners. Performance on the warehouse map is discussed in more detail in section III.

**A. Dominant Orientations Detection**

The method starts by detecting the major orientations of lines in the map. Recalling the assumption of straight line in the environment, some specific orientations are expected

\(^1\)This data set was obtained from the Robotics Data Set Repository (Radish) [18]. We thank Maxim Batalin for providing this data.
Fig. 2: Dominant orientations detection from Histogram of Oriented Gradients (HOG)

to appear more frequently in the oriented gradients, caused by long lines or collinear line segments in the map. These unidirectional features create peaks in the histograms of orientations as illustrated in Fig. 2. Due to symmetry, the angles are confined to the interval $[-\pi, \pi]$.

Histograms shown in Fig. 2 are weighted by the magnitudes of oriented gradients, therefore the strength of each gradient vector votes for dominant orientations, too. With an admissible assumption that the number of dominant directions is known, finding them from the histogram is straightforward. Dominant orientations are represented by the set $\Theta$.

$$\Theta := \{ \theta_i \mid i \in \mathbb{N}, \theta_i \in [-\pi, \pi/2] \}$$

While many environments of practical interest consist of two dominant perpendicular orientations, Fig. 2 shows the ability of this technique in handling more general scenarios.

B. Radiogram and Peak Detection

The set of dominant orientations ($\Theta$) entails the directions of long or most frequently aligned elements in the map. Projecting occupied cells of the occupancy map along the direction of $\theta_i \in \Theta$, generates signals referred to as radiograms [14]. This is demonstrated in Fig. 3a. This step is equivalent to computing a Radon Transforms [15] of the image in direction of the dominant orientations. Radiography as implemented in this work employs a filtering mechanism based on the orientation of points, rejecting if their gradients are not perpendicular to projection axes.

A set of line equations ($\mathcal{L}_{ines}$) constructs a low level spatial representation of the environment’s structure.

$$\mathcal{L}_{ines} := \{ y = m_i x + b_i \mid i \in \mathbb{N}, x, y, m_i, b_i \in \mathbb{R}\}$$

As the slope $m_i$ is already encoded in the dominant orientations $\Theta$, only the offset $b_i$ needs to be estimated. This is done via the radiogram, where peaks correspond to lines in the occupancy map. The peaks have widths that are characterised by the appearance of the entities that constitute a particular line, and this can be leveraged to influence the resolution of the resulting set of extracted lines.

Du et al. developed a technique based on continuous wavelet transform for detecting peaks in noisy signals [19], with a wavelet resembling the shape of target peaks. In this case a Ricker wavelet has been used to perform the continuous wavelet transform over the signal. Including a range of wavelets with different widths ($\sigma$) leads to robust peak detection. This technique returns only the peaks that match the characteristic of the provided wavelet restricted by the range of width. In practice, we found that $\sigma \in [1, 20]$ is a sufficient range of wavelet’s width scale for detecting peaks.

Magnitude of the peaks represent the length of the lines, interpreted as the importance of that particular set of elements. The magnitude values of these peaks can be used for adjusting the resolution of the cell decomposition in later stage, and thus the level of abstraction. It’s possible to coarsen subdivision by rejecting less important peaks. This aspect is discussed more in section III.

C. Cell Decomposition

Given the equation of lines by the set $\mathcal{L}_{ines}$ which represents structure of the environment, a polygonal cell decomposition enables us to extract higher level informations from the map. Cell decomposition in this work has a similar implementation as the polygonal decomposition explained by Kloetzer and Ghita [20]. This decomposition results in polygon faces (cells) with different shapes and sizes and is stored in a subdivision
Fig. 4: Illustration of subdivision and adjacency graph, duality and validity. Thick red lines and red dots are edges and vertices of the subdivision. Thin blue lines and blue squares are links and nodes from adjacency graph. Solid lines represent valid edges and links, while dashed lines encode invalidity.

Fig. 5: Feature space of edges from subdivision. This space illustrates possible circumstances where an edge is identified as a wall.

Duality and adjacency graph: Each edge in subdivision is neighboring two faces, which correspond to a link and two nodes from adjacency graph. Nodes in adjacency graph represents faces in subdivision and the link between two nodes encodes the connectivity between them. This duality is represented by the dual function D, where:

\[ D(n) \in \{f\}, D(f) \in \{n\}, D(l) \in \{e\}, D(e) \in \{l\} \]

Edges in subdivision have a property called cogency, which means there is a physical separation (e.g. wall) between neighboring faces. Dual to edges’ cogency, links have a property of connectivity indicating the connectivity between nodes (i.e faces). A “cogent” edge in subdivision is dual with a “disconnected” link in adjacency graph. In Fig 4b, assuming only the middle face is occupied, valid edges and links are represented by solid lines, where invalid edges and links are visualized via dashed lines.

D. Subdivision and Abstraction

Edges in subdivision could be candidates of being physical elements like walls in the map. It should be noted that here a wall is any entity perceivable by laser scanner as a straight line segment. Through the process of cogency evaluation, candidate edges are either accepted or rejected as walls. This process will lead to an abstract representation of the map. Cogency of an edge \( C(e_i) \) is evaluated upon two features, face disparity and edge occupancy \( O(e_i) \) (visualized in Fig. 5d) and is defined:

\[ \text{face disparity} = |O(f_1) - O(f_2)| \]
\[ C(e_i) = (\text{face disparity} > t_1) \vee (O(e_i) > t_2) \]

where \( (f_1, f_2) = N(e_i) \) and \( t_1, t_2 \) are thresholds. If the disparity of two neighboring faces is high, it means one being occupied and the other being free space. This implies there must be a physical element separating these two faces, hence the edge is representing a separating wall as in Fig. 5. Sometime there are walls in the middle of free space of the environment with both sides known to be free, where
Fig. 6: Subdivision encodes informations for abstraction.  

Fig. 7: The subdivision in green, and adjacency graph (only valid links) in blue. Nodes of adjacency graph are located at the center of faces and the links in this graph represent connectivity between faces.

E. Adjacency Graph and Semantic Labeling

As mentioned earlier nodes and links in adjacency graph are dual to faces and edge from subdivision. When a cogent edge represents an isolation between neighboring faces, connectivity of the dual link is defined as the complement of cogency.

\[ C(l_i) = \neg C(D(e_i)) \]

Fig. 7 illustrates that the valid links in the adjacency graph correspond to a connectivity map of the open space.

Metric attributes of the adjacency graph make it suitable for matching semantic templates that are metric-topological. Algorithm 1 presents an example for matching different types of junction patterns. In this implementation, \( T \) models two types of junctions, T-type and X-type, as illustrated in Fig. 8. It should be noticed that neither the templates in Fig. 8 nor Algorithm 1 presents a comprehensive set of semantic labels and matching algorithms. Corridors as illustrated in Fig. 8c are detected with a similar technique. The result of semantic analysis is shown in Fig. 8d.

III. RESULTS AND DISCUSSION

The method proposed in this work is tested on data from a real-world warehouse of bathroom accessories. The occupancy map (in Fig. 9a) was built using an open implementation of a SLAM algorithm based on Rao-Blackwellized particle filter called “GMapping” [21]. Performance of the method is demonstrated through semantic labeling and map abstraction (Figures 9e and 9d).

### Algorithm 1 template matching in adjacency graph: example of “junction” detection

**INPUT:** adjacency graph  
for \( n_i \in N_{nodes} \) do  
valid junction \( \leftarrow \) TRUE  
for all \((n_l, n_k) \in N(n_i), l \neq k\) do  
if \( N(n_l) \cap N(n_k) \neq \{n_i\} \) then  
valid junction \( \leftarrow \) FALSE  
end if  
end for  
\( n_i = \begin{cases} \{ T(\text{size}(N(n_i))) \} & \text{if valid junction} \\ \emptyset & \text{otherwise} \end{cases} \)  
end for  

\[ T(n) := \begin{cases} \text{T type} & n = 3 \\ \text{X type} & n = 4 \\ \emptyset & \text{otherwise} \end{cases} \]

Fig. 8: Some templates for pattern detection from adjacency graph are presented along with result of applying to a real-world map. Green lines in templates are valid connection between nodes, on the contrary red lines imply blocked access.

**Choice of Resolution:** Degree of the cell decomposition’s resolution comes from spatial characteristic of the physical entities in the environment. If the elements of the environment are distributed homogeneously with similar length, they would provoke similar peaks in radiogram signal. In reality peaks have different amplitude and width, which represent how many and how well things line up in the environment over this particular line in the model \( (L_{lines}) \). Tuning the sensitivity of peak detection technique with these two values (amplitude and width) will result in different resolutions in the cell decomposition. Figures 9b and 9c show how the change of resolution affects the decomposition. The optimum resolution depends on the application. For instance a high resolution decomposition preserving more details from the environment would be suitable for path planning (Fig. 9e).
Coarser resolution appear to be more appropriate for semantic labeling of regions. For example in Fig. 9d, corridors are marked in green, X-junctions with blue circle, T-junctions with red circles, blue lines show spatial connectivity, and red lines are the map abstractions.

**Deficiency of Semantic Labels:** Each label is the result of a possible match between some predefined templates and the adjacency graph. As it is noticeable in Fig. 8d and Fig. 9d, there exist few misclassifications of X-type and T-type junctions. This imperfection is rooted in the degeneracy of adjacency graph. A map of partially explored environment leads to some degeneracies in the adjacency graph, such as stub-like links or missing links at the frontier of unexplored areas. While these degeneracies could be minimized with a fitting choice of resolution, this would not fully eliminate them. Current work targets this shortcoming via improved graph matching and richer templates, which is beyond the scope of this paper.

**IV. CONCLUSION AND FUTURE WORK**

Robots with a heightened awareness of their surroundings are key for novel applications in automation, such as intelligent warehouses with integrative inventory management. In addition to increasing the reliability and robustness of operation, it can provide crucial background knowledge for adaptive learning or reasoning in novel situations. The presented work provides a concrete method for increasing the level of awareness beyond the state of the art.

The proposed method interprets occupancy maps to bring out underlying spatial characteristics in a format readily usable by machines and humans. The problem is approached from a global perspective by extracting dominant orientations, assuming the presence of straight lines that are important yet hard to detect. We employ radiogram analysis to fit independently oriented and spaced lines to the occupancy map. After cell decomposition over those fitted lines, two corresponding data structures capture topological characteristics and ease subsequent semantic labeling.

We present a case study of two real-world maps, one from a commonly available dataset and the other from data we collected in the warehouse of an industrial partner, to demonstrate the effectiveness of the method. Furthermore we show that a preliminary semantic labeling algorithm reliably detects corridors, and junctions.

The key features of the proposed method are the flexible and robust direction detection and the adaptive spacing of lines. At this stage, the number of orientations is given in advance. Clustering techniques could be used to generalize this aspect. No restriction is placed on the relative angle between dominant orientations. Control of the spatial resolution at which grid lines should be inferred (scale of the Ricker wavelet) influences the abstraction level of the resulting topological information. This is currently a (rather straightforward) manual process, inferred from the type of environments subject to our experiments.

Fig. 9: Cell decomposition of a real-world warehouse of bathroom accessories, with coarse and fine resolution. While coarse decomposition is more suitable for semantic labeling and extracting the layout of the environment, fine decomposition is more appropriate for detailed abstraction and connectivity maps.
Improvement of the proposed method will include a study of spatial frequency to better understand and detect features (such as pillars) that occur with some underlying regular pattern. This is expected to help with low-quality maps and enable automatic adaptation of the spatial resolution. An important next milestone in the project is the inclusion of visual perception. In [22] we proposed a modeling technique for structured environment by the means of vision sensors. That study starts with mapping a warehouse by its infrastructural landmarks (pillars). It is followed by modeling the map with a “repetitive canonical geometric-semantic” model. While the model provides an abstraction of the geometry of the environment by its landmarks and corridor boundaries, it also implies semantics on a conceptual levels. Next goal in our research would be to develop a combination technique, that is able to associate maps and their inferred knowledge from different modalities.

REFERENCES


