From images to rooms

Olaf Booij, Zoran Zivkovic and Ben Kröse
Intelligent Systems Laboratory,
University of Amsterdam, The Netherlands

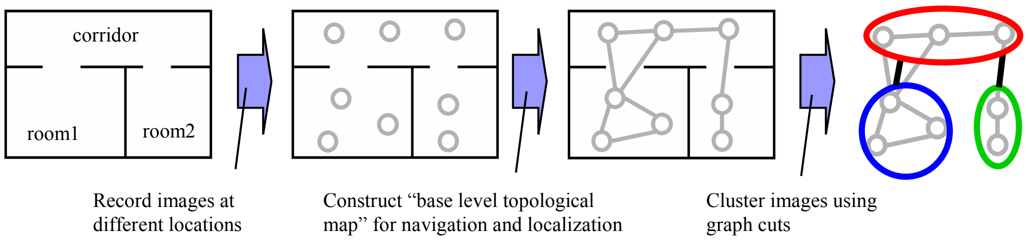

Figure 1. An overview of the algorithm.

***Abstract*— In this paper we start from a set of images obtained by the robot while it is moving around an environment. We present a method to automatically group the images into groups that correspond to convex subspaces in the environment which are related to the human concept of rooms. Pairwise similarities between the images are computed using local features extracted from the images and geometric constraints. The images with the proposed similarity measure can be seen as a graph or in a way a base level dense topological map. From this low level representation the images are groped using a graph-clustering technique which effectively finds convex spaces in the environment. The method is tested and evaluated on challenging data sets acquired in real home environments. The resulting higher level maps are compared with the maps humans made based on the same data[1].**

I. INTRODUCTION

Mobile robots need an internal representation for localization and navigation. Most current methods for map building are evaluated using error measures in the geometric domain, for example covariance ellipsis indicating uncertainty in feature location and robot location.

Now that robots are moving into public places and homes, human beings have to be taken into account. This changes the task of building a representation of the environment. Semantic information must be added to sensory data. This helps to enable a better representation (avoid aliasing problems), and makes it possible to communicate with humans about its environment. Incorporating these tasks in traditional map building methods is non trivial. Even more, evaluating such methods is hard while user studies are difficult and there is a lack of good evaluation criteria.

One of the more complicated issues is what sort of spatial concepts should be chosen. For most indoor applications, objects (and their location) and rooms seems a natural choice. Rooms are generally defined as convex spaces, in which objects reside, and which are connected to other rooms with 'gateways' [1], [2]. In [3] a hierarchical representation is used in which at the low level the nodes indicate objects, at a higher level the nodes represent 'regions' (parts of space defined by collections of objects) and at the highest level the nodes indicate 'locations' ('rooms'). However, detecting and localizing objects is not yet a trivial task.

In this paper we consider the common concept of 'rooms'. We present our appearance based method to automatically group images obtained by the robot into groups that correspond to convex subspaces in the environment which are related to the human concept of rooms. The convex subspace is defined as a part of the environment where the images from this subspace are similar to each other and not similar to the other subspaces. The method starts from a set of unlabelled images. Every image is treated as a node in a graph, where an edge between two nodes (images) is weighted according to the similarity between the images. We propose a similarity measure which considers two images similar if it is possible to perform 3D reconstruction using these two images [4], [5]. This similarity measure is closely related to the navigation task since reconstructing the relative positions between two images means also the it is possible to move the robot from the location of where one images is taken to the location where the other image is taken given that there are no obstacles in between. We propose a criterion for grouping the images from convex spaces. The criterion is formalized as a graph cut problem and we present an efficient approximate solution. In an (optional) semi-supervised paradigm, we allow the user to label some of the images. The graph similarity matrix is then modified to incorporate the user-supplied labels prior to the graph cut step.

Section II presents a short overview of the related work. Section III describes our method of constructing a low level appearance based map. In Section IV it is explained how to find parts of this map belonging to convex spaces in the environment. The method used for resampling the datasets is described in Section V. In Section VI we report the experiments we did in real home environments. Our approach is also compared to other similarity measures and standard k-means clustering. Finally we draw some conclusions and discuss future work in Section VII.

[1]The work described in this paper was conducted within the EU FP6-002020 COGNIRON ("The Cognitive Companion") project.

II. Related work

The traditional topological maps represent the environment as a graph where the nodes present distinctive locations and edges describe the transitions [1]. The distinctive locations can be obtained from the geometric map, e.g. using Voronoi graphs [6], [7] or from images, for example using fingerprint representation as in [8]. However, the extracted distinctive locations are mainly related to the robot navigation task and not the human concepts such as the rooms.

Another related task is the task of place or location recognition. To distinguish between different rooms, often visual cues are used, such as color histograms [9] or visual fingerprints [10]. A combination of spatial cues and objects detected in images taken from that room has been used by [11]. Instead of explicit object detection, also implicit visual cues such as SIFT features have been used [12]. More general problem of recognizing scenes from images is addressed [13]. However all these approaches assume that the human given labels are provided.

We present here an unsupervised algorithm to group the images into groups that are related the human concept of rooms. Our approach is similar to [5] where the images are also grouped on basis on their similarities. Similar approach was also used in [14] but for the task of of finding object categories from images. In this paper we present a grouping criterion that is more appropriate for detecting convex spaces. Furthermore, in [5] the data is obtained in a highly controlled way by taking the images at uniformly spaced locations. Here we will consider the realistic situation where the data is obtained by just moving the robot around the environment. The graph clustering will then depend on the robot movements and we propose a re sampling scheme to improve the results, see Section V. Finally, we consider a semi-supervised approach where the user provides a number of labels.

III. Image similarity measure

We start from a set of unlabelled images. In all our experiments the omnidirectional images were used taken by a mobile robot while driving through the environment (see figure 2 for the image positions of of one of the data sets used for testing). Every image is treated as a node in a graph, where an edge between two nodes (images) is weighted according to the similarity between the images. This graph can be seen as a topological map. Various similarity measure can be used. We will use here the similarity measure as in [5]. We define that there is an edge between two nodes in the graph if it is possible to perform a 3D reconstruction of the local space using visual features from the two corresponding images. We use SIFT features [15] as the automatically detected landmarks. Therefore an image can be summarized by the landmark positions and descriptions of their local appearance. The 3D reconstruction was performed using the 8 point algorithm [16] constrained to planar camera movement [17] and the RANSAC estimator was used to be robust to false matches [16]. A big advantage of such similarity measure over the pure appearance based measures is that it also considers geometry [4]. Therefore the chance is small that images from two different rooms are found similar while they might be similar in appearance [5].

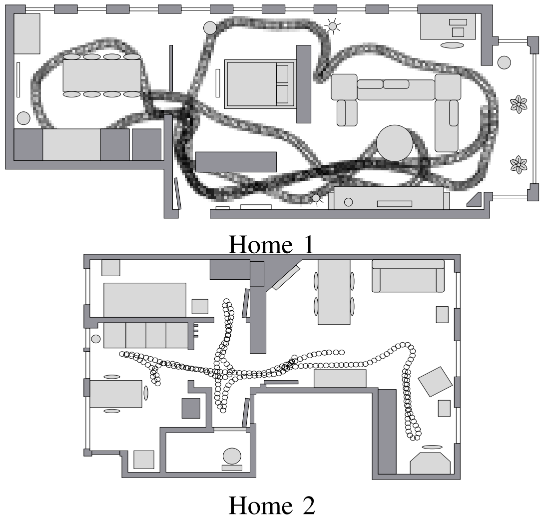

Fig. 2. Ground floor maps of the two home environments. The circles denote the positions of the robot, according to the wheel encoders, from which an image was taken.

As the result of N images we obtain a graph that is described with a set $\mathcal{S}$ of N nodes and a symmetric matrix W called the 'similarity matrix'. For each pair of nodes $i,j\epsilon[1,...,N]$ the value of the element W_{ij} from the matrix W defines similarity of the nodes. In our case this is equal to 1 if there is a link between the nodes and 0 if there is no link. Examples of such a graphs that we obtained from real data sets are given in Figure 4. If there is a non-zero edge in the graph this also means that if the robot is at one of the connected nodes (corresponding to one image), it can determine the relative location of the other node (corresponding to the other image). If there are no obstacles in between, the robot can directly navigate from one node to the other. If there are obstacles, one could rely, for example, on an additional reactive algorithm for obstacle avoidance using range sensors. In this sense the graph obtained using the proposed similarity measure can be seen as a base level dense topological map that can be used for navigation and localization.

This graph contains, in a natural way, the information about how the space in an indoor environment is separated by the walls and other barriers. Images from a convex space, for example a room, will have many connection between them and just a few connections to some images that are from another space, for example a corridor, that is connected with the room via a narrow passage, for example a door. By clustering the graph we want to obtain groups of images that belong to a convex space, for example a room.

IV. Grouping images

Starting from the graph representation we will group the images by cutting the graph $(\mathcal{S}, W)$, described above, into K separate subgraphs $\{(\mathcal{S}_1, W_1)..., (\mathcal{S}_K, W_K)\}$. If the subgraphs (clusters) correspond to convex subspaces we expect that there will be many links within each cluster and a few between the clusters. The subgraphs should also be connected graphs. This is formalized as a graph cut criterion further in this Section. An efficient approximate solution is also presented.

Note that we assume that the images are recorded at positions that approximately uniformly sample the available space. If this is not true the images from the positions close to each which are usually very similar tend to group together and the resulting clusters depend on the positions where the images are taken.

A. Grouping criterion

We will start by introducing some graph-theoretic terms. The *degree* of the i-th node of a graph $(\mathcal{S}, W)$ is defined as the sum of all the edges that start from that node: $d_i = \sum_j W_{ij}$. For nodes $\mathcal{S}_j$ (where $\mathcal{S}_j$ is a subset of $\mathcal{S}$), *volume* is defined as $vol(\mathcal{S}_j) = \sum_i d_i$. $vol(\mathcal{S}_j)$ describes the "strength" of the interconnections within the subset $\mathcal{S}_j$. A subgraph $(\mathcal{S}_j, W_j)$ can be "cut out" from the graph $(\mathcal{S}, W)$ by cutting a number of edges. The sum of the values of the edges that are cut is called a graph cut:

$$cut(\mathcal{S}_j, \mathcal{S}\backslash\mathcal{S}_j) = \sum_{i\epsilon\mathcal{S}_j, j\epsilon\mathcal{S}\backslash\mathcal{S}_j} W_{ij} \quad (1)$$

where $\mathcal{S}\backslash\mathcal{S}_j$ denotes the set of all nodes except the ones from $\mathcal{S}_j$. One may cut the base level graph into q_1 clusters by minimizing the number of cut edges:

$$\sum_j^{q^1} cut(\mathcal{S}_j, \mathcal{S}\backslash\mathcal{S}_j). \quad (2)$$

This would mean that the graph is cut at the weakly connected places, which in our case would usually correspond to natural segmentation at doors between the rooms or other narrow passages. However, such segmentation criteria often leads to undesirable results. For example, if there is an isolated node connected to the rest of the graph by only one link, then (2) will be in favor of cutting only this link. To avoid such artifacts we use a *normalized* version:

$$\sum_j^{q^1} \frac{cut(\mathcal{S}_j, \mathcal{S}\backslash\mathcal{S}_j)}{vol(\mathcal{S}_j)}. \quad (3)$$

Minimizing this criterion means cutting a minimal number of connections between the subsets but also choosing larger subsets with strong connections within the subsets. This criterion naturally groups together convex areas, like a room, and makes cuts between areas that are weakly connected.

However, the criterion (3) can lead to solutions where the clusters present disconnected graphs. The requirement that the subgraphs should also be connected graphs need to be considered also in addition.

B. Approximate solution

For completeness of the text we briefly sketch a well-behaved spectral clustering algorithm from [18] that leads to a good approximate solution of the normalized cut criteria (3):

1) Define D to be a diagonal matrix of node degrees $D_{ii} = d_i$ and construct the normalized similarity matrix $L = D^{-1/2}WD^{-1/2}$.
2) Find $x_1, ..., x_K$ the K largest eigenvectors of L and form the matrix $X = [x_1, ..., x_K] \in \mathcal{R}^{N\times K}$.
3) Renormalize rows of X to have unit length $X_{ij} \leftarrow X_{ij}/(\sum_j X_{ij}^2)^{1/2}$.
4) Treat each row of X as a point in $\mathcal{R}^K$ and cluster using for example the k-means algorithm. Instead of the k-means step in [19] a more principled but more complex approach is used, following [20] where a good initial start for the k-means clustering is proposed. We tested the mentioned algorithms, and in practice, for our type of problems, they lead to similar solutions.
5) The i-th node from $\mathcal{S}$ is assigned to cluster j if and only if the row i of the matrix X was assigned to the cluster j.

Although in practice very rarely, the normalized cut criteria (3) can lead to disconnected solutions as mentioned above. A practical split and merge solution to ensure that the subgraphs are connected is as follows:

1) group the images using the normalized cut criteria (and using the spectral clustering technique).
2) Split step: if there are disconnected subgraphs in the result generate new clusters from the disconnected subgraph components.
3) Merge step: the connected clusters that minimize the normalized cut criteria (3)should be merged.

The final result presents a practical and efficient approximate solution for our criterion from the previous section. The exact solution is a NP-hard problem and usually not feasible.

C. Semi-supervised learning

This framework allows the introduction of weak semisupervision in the form of pairwise constraints between the unlabelled images. Specifically, a user may specify cannotgroup or must-group connections between any number of pairs in the data set. Following the paradigm suggested in [21], we modify the graph $(\mathcal{S}, W)$ to incorporate this information to assist category learning: entries in the affinity matrix S are set to the maximal (diagonal) value for pairs that ought to be reinforced in the groupings, or set to zero for pairs that ought to be divided.

V. REALISTIC (NON-UNIFORM) SAMPLED DATA

The images should be recorded at positions that approximately uniformly sample the available space. However, this is often difficult to perform in practice. For example some of the data sets we will consider in the experimental section were recorded by letting the robot record the images at regular time intervals. For such data the clustering will depend on the robot movements. An illustration of a non-uniformly sampled data set is given in Figure 3. The images taken close to each other depicted in the figure near the transition from 'room2' to 'corridor' will usually be similar to each other and therefore grouped together. The on-line appearance topological mapping

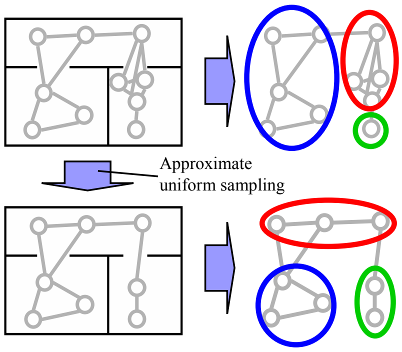

Fig. 3. Top image depicts an example of non-uniformly sampled data and undesired clustering results. The clustering results can be improved by detecting such situations and generating a new graph with approximately uniformly sampled images as depicted below.

[8] will also suffer from the same problem. In this Section we will use information about Euclidean geometric distances between the images and present a simple sampling approach aimed to approximate the uniform sampling of the space and improve the clustering results.

A. Importance sampling

Let there be N images recorded while robot was moving around the environment and let $x^{(i)}$ denote the 2D position where i-th image was recorded. We can consider $x^{(i)}$-s as N independent samples from some distribution q. A sample based approximation is $q(x) \approx \sum_{i=1}^{N} \delta(x - x^{(i)})/N$. Then we can approximate uniform distribution using importance sampling:

$$Uniform(x) = c = \frac{c}{q(x)} q(x) \approx \sum_{i=1}^{N} \tilde{w}^{(i)} \delta(x - x^{(i)}) \quad (4)$$

where $\tilde{w}^{(i)} = w^{(i)} / \sum_{j=1}^{N} w^{(j)}$ and $w^{(i)} = c/q(x^{(i)})$. One can interpret the $\tilde{w}^{(i)}$ as correction factors to compensate for the fact that we have sampled from the "incorrect" distribution $q(x)$. Approximate uniform sample can be generated now by sampling from the sample based approximation above. This is equivalent to sampling from the multinomial distribution with coefficients $\tilde{w}^{(i)}$. The original distribution of the original sampling $q(x^{(i)})$ can be estimated for example using a simple K-nearest neighbor density estimate $q(x^{(i)}) \sim 1/V$ where the $V = d_k(x^{(i)})^2$ and $d_k(x^{(i)})$ is the distance to the k-th nearest neighbor in the Euclidean 2D space. The distances can be obtained form odometry or some SLAM procedure. Alternatively the distances can be approximated from the images directly. For all our data we used the $k = 7$.

B. Practical algorithm

We start with the original graph $(\mathcal{S}, W)$ and an empty graph $(\mathcal{S}^{resampled}, W^{resampled})$. The practical algorithm we will be using is as follows:

1) Compute the local density estimates and the weight factors $\tilde{w}^{(i)}$.
2) Construct a new graph sampling N samples from the multinomial distribution with coefficients $\tilde{w}^{(i)}$. The corresponding nodes and links from the original graph $(\mathcal{S}, W)$ are added to the new graph $(\mathcal{S}^{resampled}, W^{resampled})$.
3) if the new graph $(\mathcal{S}^{resampled}, W^{resampled})$ is not connected continue sampling and adding nodes as in the previous step until it gets connected.

The result is the new graph $(\mathcal{S}^{resampled}, W^{resampled})$ where the images come from positions that approximately uniformly sample the available space.

VI. Experiments

The method of finding the convex spaces in an environment is tested in two real home environments and is compared to the annotation based on the same sensor data. Our mobile robot was driven around while taking panoramic images with an omnidirectional camera, see figures 2 for ground floor maps of the environments and the positions where images were taken. The task of building a map using these image sets is challenging in a number of ways. First of all the lighting conditions were not good, much worse than the conditions during previous evaluations in office environments. Also, people were walking through the environment blocking the view of the robot. Furthermore, the robot was driven rather randomly through the rooms, which has the effect that some parts of the environment are represented by a lot of images while others parts only with a few (see www2.science.uva.nl/sites/cogniron/ for videos acquired by the robot).

The data sets were annotated by a inexperienced person, based solely on the sensor data and the maps as shown in figures 2 but without the robot positions. The person had never visited one of the two houses. For both homes labels were provided corresponding to the rooms, from which one should be picked per panoramic image. Between some of the rooms there was no good geometrical boundary separating them, so from most places in one room the other room was still clearly visible and vise verse. This is common in real home environments but makes conceptualization of it harder.

From both image sets an appearance graph is made using the methods explained in III. These graphs are then used as input for the clustering algorithm to find convex spaces in environment, first with all images and then with a subset obtained by resampling. The results are compared with the annotation, to see how well the convex spaces found by clustering correspond to separate rooms.

A. Results

In Figure 4 it can be seen that the appearance based methods were quite successful in creating a low level topological map. All links of the graphs connect nodes originating from images that were taken close to each other in world coordinates. In some parts of the graph the nodes are more densely connected than others. This could be the result of bad image quality for example caused by changing lighting conditions, but it

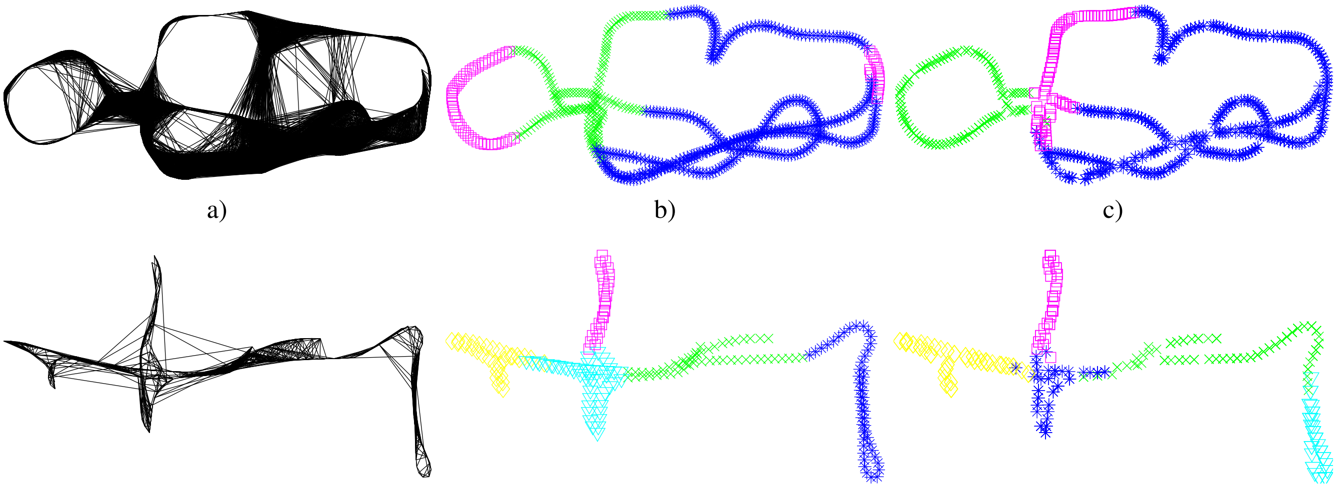

Fig. 4. The clustering results for the Home 1 (above) and the Home 2 (below) data sets. a) The appearance based graph. Each line indicates two matching images. b) The clusters found in the whole dataset. c) Clusters found in the resampled dataset. Note that the odometry data used to draw these figures are not used only for the resampling.

could also be the result of lack of features in that part of the environment.

Clustering without resampling (see Figures 4b) results in a grouping of the images which is not perfect. As can be seen in Figure 4 some of the images of Home 1 are grouped together which were taken from completely different positions and that images taken in the kitchen are split among two clusters. In Figure 4 it can be clearly seen that some images taken in the living room are grouped with images taken in the work room. After the split and merge steps these images are regrouped with the living room images.

Better clustering results are obtained after resampling the data as indicated by figure 4c. Both data sets are clustered almost perfectly, often cutting the graph at nodes corresponding to images taken at the doorpost between the rooms. The only error left is at the bedroom of Home 1, from which images are grouped with images from the living room. This is probably caused by the large opening between the two rooms, as can be seen in Figure 2.

The mismatch between the clusters found by our method and the labels provided by the annotator is made clear by the confusion matrices, see tables I to IV. Of course the clustered data does not provide a label. Each cluster is appointed the label corresponding to the true set with which it has the largest overlap, taking care that no two clusters get the same label. The percentage of correctly clustered images from home 1 is 85% for the whole dataset and 92% for the resampled set. For home 2 this was 73% and 83%.

B. Comparison with other clustering methods and similarity measures

We compare our method with the common k-means clustering and a PCA based similarity measure [22]. We used 10 PCA components and clustered the images using k-means. We also used the Euclidean distances in the PCA space and

TABLE I

HOME 1 WHOLE DATASET

True label	**Inferred label**		
	Living r	Bedroom	Kitchen
Living room	**0.9681**	0.0319	0
Bedroom	0.1832	**0.8168**	0
Kitchen	0	0.5000	**0.5000**

TABLE II

HOME 1 RESAMPLED AVERAGED OVER 10 TRIALS

True label	**Inferred label**		
	Living r	Bedroom	Kitchen
Living room	**1.0000**	0	0
Bedroom	0.3014	**0.6915**	0.0071
Kitchen	0	0.0396	**0.9604**

TABLE III

HOME 2 WHOLE DATASET

True label	**Inferred label**				
	Corridor	Living r	Bedroom	Kitchen	Work r
Corridor	**0.6812**	0.1159	0.2029	0	0
Living room	0	**0.5732**	0	0	0.4268
Bedroom	0	0	**1.0000**	0	0
Kitchen	0.0556	0	0	**0.9444**	0
Work room	0	0	0	0	**1.0000**

TABLE IV

HOME 2 RESAMPLED

True label	**Inferred label**				
	Corridor	Living r	Bedroom	Kitchen	Work r
Corridor	**0.6344**	0.0323	0.2473	0.0860	0
Living room	0.0291	**0.8301**	0	0	0.1408
Bedroom	0	0	**1.0000**	0	0
Kitchen	0	0	0	**1.0000**	0
Work room	0	0	0	0	**1.0000**

TABLE V

CLUSTERING ACCURACY FOR VARIOUS CLUSTERING METHODS FOR THE HOME 2 DATA SET. PCA PROJECTION WITH 10 COMPONENTS IS USED

PCA + k-means	PCA + spectral clustering	our method	our method (with resampling)
0.60	0.38	0.73	**0.83**

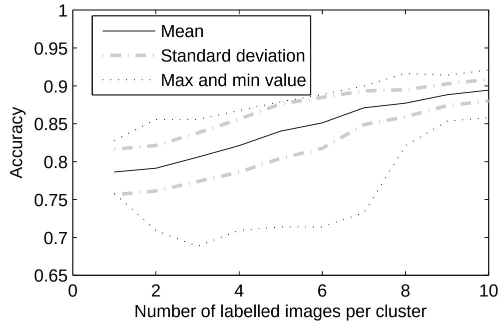

Fig. 5. The clustering accuracy for the semi supervised case - average from 100 trials. Different number of randomly chosen ground truth labels per cluster are used to simulate user input.

applied spectral clustering. The results were poor compared to our method. The results also show that this simple appearance based similarity is not suitable for spectral clustering methods.

C. Semi-supervised clustering

To demonstrate the semi supervised learning we used a set labelled points to enforce that the points with the same label should group together and the points with different labels should not group together. The set of the labelled points is randomly chosen and the results for the Home 2 data set are presented in Figure 5. The graphs show how the accuracy increases with the amount of labelled images.

VII. CONCLUSION

The experiments show that the proposed clustering method seems appropriate for finding the convex spaces by grouping images obtained by the robot. The cuts made in the graphs are at or close to the doorways dividing two rooms. The convex spaces thus found in the real home environments correspond to the concept of "room" as shown by comparing it with annotated data.

For some cuts the clustering relies on a good sampling of the data, which was clearly visible tests in Home 1. In table I it can be seen that the kitchen is split into two parts. After resampling (table I) 96% of the image annotated as the kitchen fell in a single cluster.

The proposed methods are very suitable as a basis for human robot communication about the spaces the robot travels through. The system will be developed further in this direction, with the goal to enable a robot to build a higher level map by listening to and asking a human guide. Our method naturally allows semi-supervised learning as we demonstrated, using the input of the guide. If the guide says to the robot that they just entered the kitchen, then this information should be used to build the higher level map. Problems might occur if the user is using different labels for the same space or when the clustering obtained by the robot does not correspond to the human concept. These problems need to be addressed and resolved for example through dialog with the user. Finally, the algorithms should work online in order to facilitate interaction between the map building process and the guide.

REFERENCES

[1] B. Kuipers, "The spatial semantic hierarchy," *Artif. Intell.*, vol. 119, no. 1-2, pp. 191–233, 2000.

[2] D. Kortenkamp and T. Weymouth, "Topological mapping for mobile robots using a combination of sonar and vision sensing," in *In Proc. of the Twelfth National Conference om Artificial Intelligence*, 1994.

[3] A. Tapus, S. Vasudevan, and R. Siegwart, "Towards a multilevel cognitive probabilistic representation of space," *In Proc. of the International Conference on Human Vision and Electronic Imaging X, part of the IST-SPIE Symposium on Electronic Imaging*, 2005.

[4] F. Schaffalitzky and A. Zisserman, "Multi-view matching for unordered image sets, or "How do I organize my holiday snaps?"," in *Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark*, vol. 1. Springer-Verlag, 2002, pp. 414–431.

[5] Z. Zivkovic, B. Bakker, and B. Kröse, "Hierarchical map building using visual landmarks and geometric constraints," in *Intl. Conf. on Intelligent Robotics and Systems*. Edmundton, Canada: IEEE/JRS, August 2005.

[6] H. Choset and K. Nagatani, "Topological simultaneous localisation and mapping: Towards exact localisation without explicit localisation," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 2, pp. 125–137, April 2001.

[7] P.Beeson, N.K.Jong, and B.Kupiers, "Towards autonomous place detection using the extended voronoi graph," *In Proceedings of the IEEE International Conference on Robotics and Automation*, 2005.

[8] A. Tapus and R. Siegwart, "Incremental robot mapping with fingerprints of places," in *IROS*, 2005.

[9] I. Ulrich and I. Nourbakhsh, "Appearance-based place recognition for topological localization," in *Proceedings of ICRA 2000*, vol. 2, April 2000, pp. 1023 – 1029.

[10] A. Tapus and R. Siegwart, "A cognitive modeling of space using fingerprints of places for mobile robot navigation," in *ICRA*, 2006.

[11] A. Rottmann, O. Martínez Mozos, C. Stachniss, and W. Burgard, "Place classification of indoor environments with mobile robots using boosting," in *AAAI*, 2005.

[12] F. Li and J. Kosecka, "Probabilistic location recognition using reduced feature set," in *IEEE International Conference on Robotics and Automation*, 2006.

[13] A. Torralba, K. Murphy, W. Freeman, and M. Rubin, "Context-based vision system for place and object recognition," *In Proc. of the Intl. Conf. on Computer Vision*, 2003.

[14] K. Grauman and T. Darrell, "Unsupervised learning of categories from sets of partially matching image features," *cvpr*, vol. 1, pp. 19–25, 2006.

[15] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[16] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision, secon edition*. Cambridge University Press, 2003.

[17] M. Brooks, L. de Agapito, D. Huynh, and L. Baumela, "Towards robust metric reconstruction via a dynamic uncalibrated stereo head," 1998. [Online]. Available: citeseer.csail.mit.edu/brooks98towards.html

[18] A. Ng, M. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," *In Proc. Advances in Neural Information Processing Systems 14*, 2001.

[19] S. X. Yu and J. Shi, "Multiclass spectral clustering," *In Proc. International Conference on Computer Vision*, pp. 11–17, 2003.

[20] L. Zelnik-Manor and P. Perona, "Self-tuning spectral clustering," *In Proc. Advances in Neural Information Processing Systems*, 2004.

[21] S. Kamvar, D. Klein, and C. Manning, "Spectral learning," *In Proc.of the International Conference on Artificial Intelligence*, 2003.

[22] B. Krose, N. Vlassis, R. Bunschoten, and Y. Motomura, "A probabilistic model for appearance-based robot localization," *Image and Vision Computing*, vol. 6, no. 19, pp. 381–391, 2001.