The 2010 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 18-22, 2010, Taipei, Taiwan

Building Efficient Topological Maps for Mobile Robot Localization: An Evaluation Study on COLD Benchmarking Database

Sherine Rady, *Student Member, IEEE*, Achim Wagner, and Essameddin Badreddin

Abstract—**Topological localization is a qualitative solution approach that can assist obtaining a faster quantitative metric solution by limiting the searchable space. Consequently, its efficiency is an essential requirement in hierarchical localization frameworks. This paper presents a topological map generation method with a localization scheme. Good compromise of performance measures – accuracy, memory and processing time – indicates the method's efficiency. The suggested implementations rely on information-theoretic selection of local features for node distinctive representation, and a visual codebook for compression. Testing the proposed approach on the COLD database, a recent specific benchmarking database for robotic topological mapping and localization, reveals its customization according to the vision sensor and environment characteristics. The approach guarantees over 90% localization accuracy with more than 50% overhead reduction, and is suitable for application in highly unstructured cluttered environments that are influenced by dynamics and illumination variations.**

I. Introduction

MAP building and self localization are fundamental problems in mobile robots research. Building a reliable environment model is crucial for efficient localization. According to [1,2], the difficulty of map building arises from: (i) Size of the environment; (ii) Noisy perception and actuation; (iii) Perceptual ambiguity due to environment places similarities; (iv) Cycles mapping and loop closing; (v) Unstructuredness and dynamics of environments.

Traditionally, maps are classified into metric and topological. They are further classified according to the way they are indexed, being location-based or feature-based. Metric maps are volumetric, in the sense that they offer a label for every possible location in the environment (e.g. occupancy grids). Topological maps are graph-like structures that define the environment as set of significant places with interconnections between them. Grid maps have the advantage of providing fine metric pose estimation, but on the account of extra overhead in storage, computation and maintenance requirements.

On another side, most of the current metric localization approaches work probabilistically, fusing motion models with visual or range observations [3]. Recently, hybrid maps [4,5] and integrating topological information [6] have been introduced as hierarchical structures. Initial topological node localization shrinks the metric search space either to a single space or a distribution over it, providing better memory and time management suited for large environments. Obviously, reliable topological localization is critically required within those hierarchical frameworks. Visual scene and object recognition algorithms are often used to find a solution as a place recognition problem. Such solution should be robust against perceptual aliasing and scene variability. Besides the global localization, efficient place recognition solves important related issues, such as loop closing [7] and lost or kidnapped robot situations [3]. Another important aspect is that it introduces context and semantics into the system.

The goal of this work is to build reliable and efficient topological maps, with minimal but relevant information to support hierarchical localization. Efficiency implies that the map information is robust for recognition and discriminative for accurate place identification. The small size of the map assists faster performance and real-time implementation adequacy. A realization for those objectives was achieved through a system that has been developed and tested on long terms in the indoor office environment of Automation Lab at Heidelberg University [8,9]. The system depends on highly-selected distinctive features acquired from an information-theoretic evaluation, and creates a codebook for features compression. Using a single perspective camera, the system showed high performance with 96% localization accuracy and more than 90% reduction for the extracted features.

In this paper, the proposed method is further tested on a recently-constructed standard benchmark database (COLD), primarily targeted for robot topological localization and mapping. Omnidirectional sensor images subject to varying conditions of illumination and scene dynamics are used. The benchmarking study is relevant for the evaluation and validation of the proposed method. Comparisons between the two databases are highlighted, and the efficiency of the proposed method is explicitly presented in terms of localization accuracy, map storage and computational costs.

The paper continues next with a review on the related work in section 2, the proposed method and its components description in section 3, an overview over the COLD database in section 4, the experimental work in section 5, and finally with discussion and conclusions in section 6.

II. Related Work

Visual place recognition emerged as a useful tool for obtaining topological information about the robot's position.

Manuscript received March 10, 2009. Authors are with the Automation Laboratory of Computer Engineering Institute, University of Heidelberg, 68131-Mannheim, Germany. (Phone/fax: 0049-621181-2471/2739; e-mail: {sherine.rady, achim.wagner, badreddin}@ziti.uni-heidelberg.de).

Several techniques have been proposed using panoramic vision [10,11], perspective vision [12], and sometimes combined with range sensors [13,14]. Panoramic cameras gained much popularity recently, as they provide larger field of view and create features that are invariant to the robot's orientation. For place characterization, some of the applied techniques use global features such as color histograms [15,16], sequential composite lists [13], eigenspace representation [17], and frequency filter responses [10]. Other techniques are landmark-based or employ invariant local features such as [6,11]. Other than recognizing specific place instances, the semantic categories of places (e.g. room, doorway, corridor, kitchen, office, etc.) are recognized in [14,18] relying primarily on the use of classifiers.

A highly important issue in image characterization is the quality and the size of the descriptors, especially for local descriptors. They can generate more than 1000 features per image, depending on how much detail an image contains [19]. This large size imposes a burden on the localization system, and may affect successive behaviors like tracking. Most of the current studies apply feature extraction, assuming that features are highly qualitative only, without paying attention to their size and their information content. Only few trials considered feature pruning [20], or features compression through visual vocabulary [9,21].

Additionally, robustness to environment dynamic changes has taken little consideration [12,18]. From the point of view of vision-based solutions, it is important to provide robustness against illumination variations and changes introduced by human activity (people appearing in the rooms, objects and furniture being relocated or removed), which influence the appearance of places over time.

The above two factors are regarded as a particular requirement in the context of qualitative mapping and localization, and are the main considerations of this paper.

III. Proposed Method

Our map building problem can be formulated as follows: *Given:* A mobile robot equipped with a single vision sensor, capturing image sequences I from a defined place set N of size n; A domain of features F extracted for N through a function $g(I)$; *Required:* Construct an undirected graph topological map in the form of $T{:=}(N,C)$, where C is a set of pairs indicating a spatial interconnection between node N_i and node N_j; $i,j=1,\dots,n$; and $i \neq j$, with the minimal set $f_i \subset F$, that maximizes the node recognition probability $P(N_i \mid f_i)$; *Constraints:* No a priori environment specification (objects, landmarks); environment is influenced by scene dynamics and varying illumination. The solution to localization is to find the robot's most probable position p; $p \in N$, using T.

Fig. 1 introduces the idea behind the map generation and its realization for topological localization. The main modules are the feature evaluation and feature compression. They are applied on high-dimensional invariant local descriptors. These modules achieve a double benefit; the selection of best place discriminative features, and the reduction in the large number of features. Hence, they guarantee generating a small size relevant map. The two modules are calculated offline. A training dataset is required for the features evaluation. The output of the features evaluation is a set of feature clusters, which are fed into a codebook for compression. The collected dataset is often subject to dynamic changes, since the visual appearance in indoor environments varies vastly (day and night, artificial light on and off, objects moved from their places). The heterogeneity of the dataset should increase the recognition robustness.

Localization is modeled in the form of an image retrieval system, in which the robot's location is identified by comparing the features extracted from the current scene to those of the previously captured place imageset stored beforehand in the database (map). The nearest match or matches identify the robot's current position. The different components of the proposed architecture, with suggested implementations are explained in the following subsections.

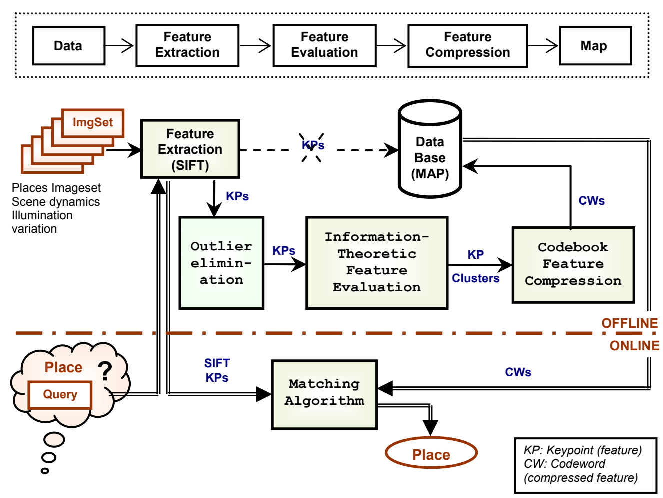

Fig. 1. Map generation idea with its realization for topological localization.

A. *Feature Extraction and Matching*

The society of computer vision provides a wealthy set of algorithms for feature detection and description. Opposed to global image signatures, local features require excessive computation time. They are, however, more robust to illumination, clutter and occlusion, which are common indoor environment properties.

Images are represented by distinctive features extracted using Scale Invariant Feature Transform (SIFT) [19]. The transform combines a scale invariant feature detector and a gradient distribution descriptor. The detector is based on Difference-of-Gaussians (DOG) acquired in successive scale space decompositions. Interest points' locations (keypoints) are identified and described by gradient histograms within a 16x16 local neighborhood. For each keypoint, a 128-dimensional vector is generated relative to its scale. The descriptor is invariant to scale and rotation, and slightly affected by illumination change, noise and small distortions.

Despite the descriptor's richness and robustness, its high dimensionality and huge size are an obvious disadvantage.

Matching is performed using a cosine distance and a voting scheme. A keypoint selective matching by imposing a threshold provides better matching for the recognition [19].

B. Outliers Detection and Elimination

Outliers exist in every data gathering process. The goal of this module is to eliminate fraud features (e.g. unstable scene features that appear temporarily, surfaces reflectance due to light), which reduce feature quality, and which may lead to potential misrecognition.

To detect outliers, a clustering technique using *k*-means is applied. Normally, inlier data belong to large and dense clusters, while outliers either do not belong to any cluster or form very small clusters [22]. The assumption holds true, since objects are normally composite and images contain lots of redundant features. Besides, stable features will appear in the image set that belong to the same scene. For this, keypoints associated with images that belong to the same place are clustered, and the extremely small clusters are removed. Though the used image matching criterion is robust to a big extent to outliers, it is recommended to preserve only consistent features in the map.

C. Information-Theoretic Features Evaluation

In order to increase the localization accuracy, places are marked, as far as possible, in a unique separable way. In other words, each place is characterized by information that makes it recognizable and distinguishable from all other places. Perceptual aliasing is reduced this way, making localization more reliable. Hence, an information-theoretic approach using an entropy measure is applied to evaluate the features with respect to their utility for categorization. The output of this evaluation is which keypoint sets give the best performance and how many approximately are sufficient.

A training dataset consisting of {*Keypoint–Class*} tuples is constructed, from which the conditional entropy of a *Class* given a *Feature Category* is calculated. The Class information is encoded in terms of higher level representation by clustering keypoints of images that belong to the same place, similarly as in the outlier detection. The extraneous redundant features that are mostly present in rich scenes, and which may refer to an exact, a part of or a characteristic appearance of an object, augment this processing. The clustered features, referenced as 'keypoint clusters' in fig. 1, will easily form the codewords of the visual codebook later. The Feature Category is the true realization of the keypoints variation, and is obtained through a histogram by sampling the entire keypoints in the dataset. *K*-means algorithm is used for clustering both the Class information and the Feature Category.

The entropy of recognizing a Class *O*, given a sampled Feature Category f_i is calculated through the posteriors:

$$H(O|f_i) = -\sum_{k} P(o_k|f_i).\log_2 P(o_k|f_i) \qquad (1)$$

for every k=1,…,Ω, where Ω is the number of instantiations of the classes, and i=1,…,Ψ, where Ψ is the number of instantiations of the feature categories.

Equation (1) distinguishes the quality of keypoints. Those keypoints that tend to appear equally likely among the different classes and image categories are less informative, and encounter high entropy values. On the opposite side, the keypoints whose occurrence is bound to few classes and image categories deliver more information (distinguishable in terms of categorization) and encounter low entropy values. Accordingly, a decision regarding each keypoint (Feature Category), whether it is useful for categorization or not, can be taken based on the information that it delivers (i.e. if entropy is low).

D. Visual Codebook Generation

Since the main objective is to obtain a small size relevant feature map, a visual codebook (CB) [23] is applied to compress the domain of the selected low-entropy clusters obtained in section 3.3. A codeword entry in the CB is represented by the centroid of each keypoint cluster, with a reference to the related images. In the localization process, no clustering is needed. Only the extracted keypoints are matched against the codewords of the CB. Since a CB preserves only one value on behalf of the keypoint cluster elements, the size of the preserved map features is again much more reduced than the low-entropy feature map.

IV. COLD Database Description

COLD (COsy Localization Database) [24] is a large-scale, testing environment for evaluating vision-based localization systems aiming to work on mobile platforms in realistic settings. The database provides a large versatile set of image sequences acquired at 3 different laboratory environments in European cities (Saarbruecken, Freiburg and Ljubljana), and under different natural variations. Similar place categories or room functionalities are found in all of the three databases.

Perspective and omnidirectional video sequences were recorded using three different mobile robot platforms. Laser range scans and odometry data were also captured for most of the sequences. In each laboratory, data was acquired in all rooms using the same camera setup. The acquisition process is done under different weather and illumination conditions (cloudy, sunny, night), and across a time span of 2-3 days. Dynamic changes like people wandering around and missing or newly added objects were introduced. Since the COLD database provides an ideal and flexible testbed for assessing localization and recognition algorithms robustness, we use an exemplar part to validate our proposed method in the next section. Example images are shown in fig. 2.

V. Experimental Results

COLD-Saarbruecken omnidirectional extended sequences B are used for experimentation. In each video, the robot navigates in 5 different functional areas in about 34 seconds;

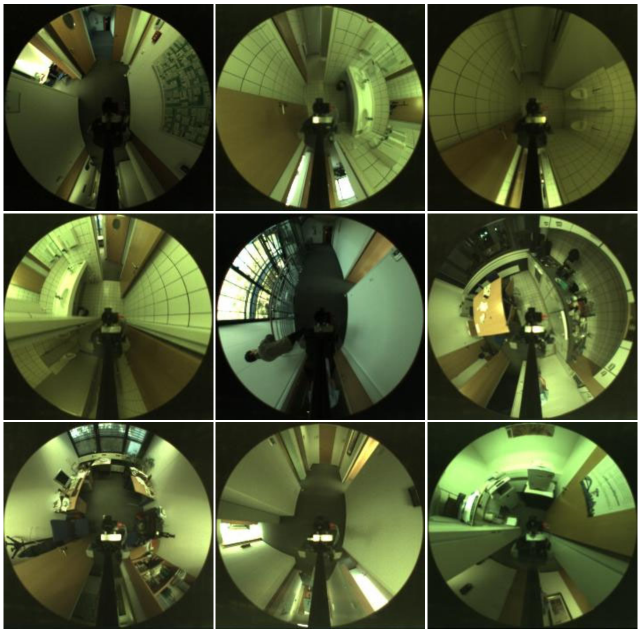

Fig. 2. Nine-place example for 5 functional categories in Saarbruecken-COLD Database (corridor, bathroom, kitchen, office room & printer area).

corridor, 1-person office, kitchen, bathroom and printer area. A topological map from 9 different places among those categories is constructed (fig. 2). The resolution of an omnidirectional image is 640x480 pixels, which accounted for a 1003x199 slightly distorted image when unwrapped.

Each place is represented by 3-9 images according to the existing amount of variation due to scene dynamics, lighting conditions, or different robot pose. For feature evaluation, a database is constructed with 49 images acquired from 3 videos with the 3 different categories (cloudy, sunny and night). A second database is constructed from 6 videos with another intensity variation for the performance evaluation. The test database has 52 images that are subject to severe dynamic variations (darkness, shadows, viewpoint change).

Applying the SIFT feature extraction, the average number of keypoints extracted per image is 333 keypoints, each with 128 dimensions. Assuming that a single number occupies 8 bytes for storage, the map size becomes 15.92 Mbytes. As mentioned in section 3.1, the cosine distance is used for matching together with a threshold to discard apart distanced matches. The threshold value is set to 60% in keypoint to keypoint matching, and to 50% in keypoint to codeword matching. An identified place is determined by the overall keypoint or codeword majority votes. As a retrieval system, the accuracy of the localization was adopted in [8] to be the ratio of correctly identified place images, with insurance that 60% of the correct images that reside in the database are retrieved (i.e. Precision measured at a Recall value of 60%).

In the clustering-based outlier detection, 9% of the keypoints are eliminated after discarding small clusters (≤ 5). To avoid discarding too many features from the images of locations with few detected features, the value of k is set to be a function of both the number of images per place and the average number of extracted keypoints per image. The filtered feature set had identical Precision-Recall performance behavior as the original features extracted by SIFT. This feature set is used for training in the next study.

Different values for the real keypoint cluster variation parameter Ψ are investigated for the information-theoretic evaluation explained in section 3.3. Values between 100 and 2000 are tested for a total of 14848 keypoints, and the consequent effect of eliminating keypoints of relatively high-entropy values is monitored. Fig. 3.(a-c) shows the measured Precision versus different high-entropy keypoints elimination percentages for Ψ = 100, 500 and 1512. The graphs maintain almost a constant Precision before it starts decreasing at 60% high-entropy features elimination. The 500 cluster variation is considered ideal since it undergoes a smooth variation.

Fig. 3.(d-f) shows the relation between Precision and Recall as a performance behavior for the retrieval process, and as a function of the different elimination percentages in fig. 3.(a-c). In order to demonstrate the effect of the feature evaluation, performances are also compared to original SIFT features after outlier detection (red curve). The relationships exhibit a dense bundle of curves, in which performance is close to and sometimes better than SIFT. Such bundle corresponds to the constant Precision level indicated in fig. 3.(a-c). This means that localization accuracy is almost like SIFT. The green curve is the one selected for the codebook generation. It has a filtering ratio of 52% for high-entropy features, and in other saying a 48% low-entropy feature set.

Fig. 4 shows the localization performance of the 48% low-entropy features for Ψ = 100, 500, 1000, 1512, 2000 for the test database. The map contains on average 6886 keypoints. Low cluster variations (100, 500) show better performances in comparison with the features extracted by original SIFT and after outlier detection (OD), as well as compared to the higher cluster variations which is an advantage. It is worth mentioning that the recognition is higher in the test database, with a value of 93% for the low-entropy features (Ψ=100 and 500) versus 86% for SIFT.

Fig. 5 shows the performance of five CB examples generated from the previous entropy-based feature sets. The map contains only 575 keypoints, equivalent to 12 times compression than the entropy-based feature map, and 28 times than the SIFT map. Matching using the CB has much lower localization accuracy than the entropy-based features, as well as than SIFT (57-64%). In this case, the CB fails to achieve the expected high localization accuracy.

Table (1) summarizes the localization performance of working modules versus different criteria. Measurements are obtained in Matlab/Windows environment on a P4, 3.2 GHz PC, with the databases residing on external disk storage, and are calculated as the average of the five studies carried out on Ψ. As indicated, an average reduction in the map size of 57% is obtained for the qualitative entropy-based features, and with better accuracy than the original SIFT features

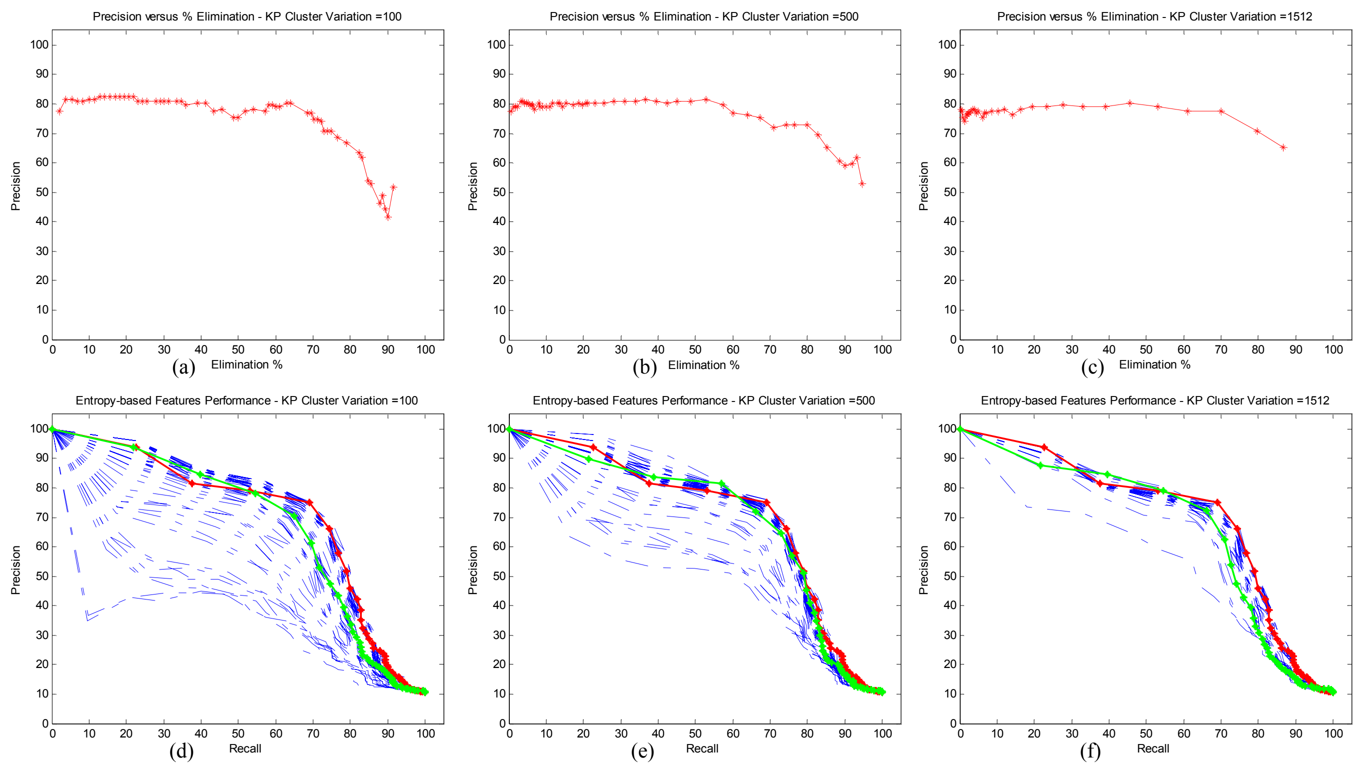

Fig. 3: (a-c) Average localization (Precision) versus different percentages of high-entropy features elimination (Histogram size for keypoint cluster variation = 100, 500 and 1512). (d-f) Average Precision-Recall performance for the different elimination percentages of 3.(a-c). The red curve is the performance of SIFT algorithm (reference). The green curve shows features elimination percentage of 52%, chosen for codebook generation.

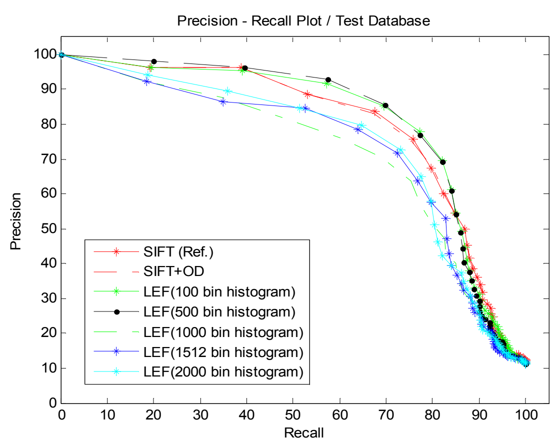

Fig. 4. Localization performance based on 48% low-entropy features.

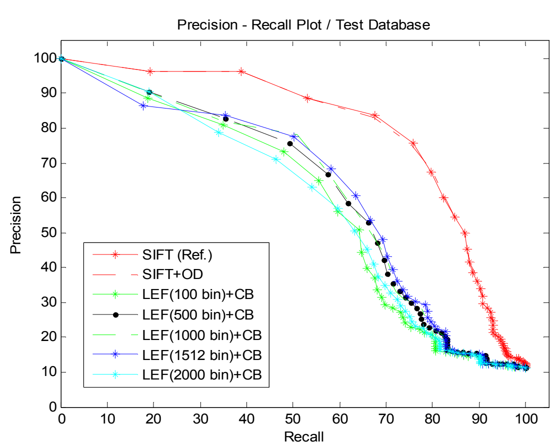

Fig. 5. Localization performance for codebooks generated from 48% low-entropy features.

TABLE I. AVERAGE SYSTEM PERFORMANCE – MODULE VERSUS CRITERIA

Module / Parameter	Original SIFT Alg.	Original SIFT + outliers detection	Entropy -based Features	Code-book
Map Size(Mbytes)	15.92	14.48	6.72	0.56
Avg. KPs per image	333	302	140	12
Precision at 60% Recall (%)	86.19	85.93	90.84	61.95
Precision at 20% Recall (%)	96.15	96.15	97.09	89.86
Localization time (sec)	3.582	3.2271	2.2388	1.5440
Memory reduction (%)	-	9.07	57.77	96.47
Localization time reduction (%)	-	3.9	33.33	54.02

(90% for entropy-based features versus 86% for SIFT). The codebook records 96% reduction, but at a lower accuracy. Localization time is also reduced for entropy-based features and codebook by 33% and 54% respectively compared to SIFT. This timing includes 1.271 seconds, which is the SIFT execution time on the specified machine. Excluding this timing, reduction in matching and retrieval times is linear with the storage reduction as presented in [8].

VI. Discussion & Conclusion

This work is concerned with generating efficient maps with the smallest set of relevant features that support discriminative place categorization for localization purposes. A method is proposed based on information-theoretic evaluation and selection of environment natural features and attaching them to a topological map in compressed format. An implementation for the method was tested before in the indoor environment of Automation Laboratory at Heidelberg University [8,9], where a perspective camera of resolution 640x480 is used. Wide-view images are obtained by the stitching of sequential images procedure, which generated 1500x300 pixel images. The Heidelberg database showed very good performance using the proposed approach (96% accuracy with 90% reduction for the extracted features).

In comparison to the COLD images, the increase in the visual field of omnidirectional camera comes at the cost of resolution. The image set contains less detail, not only due to the resolution, but also because the environment itself is less cluttered (e.g. large wall areas). The average number of features extracted per scene reveals this fact (333 for the COLD database versus 1000 for the Heidelberg database).

Experimentations on the COLD database were successful too, but showed different performance. It was found that the selected features through the entropy measure outperform the original extracted SIFT features, recording a localization accuracy of 90.8%. It managed as well to reduce more than 50% of the map size, and 30% of the localization time. The codebook module, however, didn't provide high localization accuracy. This is because the images undergo severe illumination changes and contain less detail, a matter that only 12 features per image are insufficient to efficiently categorize the scene. Furthermore, the variation between the codewords was not that high, which led to misrecognition.

It is important to mention that the choice of the number of clusters in clustering approaches is an important factor. This prevents the data from undergoing little or extra division, and hence misses or loses meaningful information content.

Within our experimentations, the localization is judged as a retrieval system with the localization accuracy set to the Precision at 60% Recall. This puts a restriction that more than half of the images bound to the best match should be retrieved. This setting can be used to derive a probability distribution over places to be fused with filters for speeding up metric localization [3]. It is still also appropriate to set the accuracy to lower Recall values, and obtain a single place solution (best match), in case the recognition rate is high.

The same tested database is assessed in [18] using Harris-Laplace, SIFT and support vector machines. The authors report a recognition rate of 88% for the case that the system is trained and tested with the same weather conditions.

In conclusion, a customized method using either feature evaluation and compression components, or using solely the feature evaluation component, provides substantial map size and time savings besides accuracy for localization purposes. The codebook performance is function of size and variation of the extracted features (i.e. environmental details). Within hierarchical localization frameworks, which suit large-scale environments, small size efficient maps can be constructed, and a reliable topological localization scheme can be safely integrated. This is our future work.

REFERENCES

[1] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics.* MIT Press, Cambridge, MA, 2005, pp. 281-284.

[2] G.N Desouza and A.C Kak, "Vision for mobile robot navigation: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(2), pp. 237-267, 2002.

[3] S. Thrun, D. Fox, W. Burgard, and F. Dellaert, "Robust Monte Carlo localization for mobile robots," *Artificial Intelligence Journal*, 128(1-2), pp. 99-141, 2001.

[4] N. Tomatis, I. Nourbakhsh, and R. Siegwart, "Hybrid simultaneous localization and map building: a natural integration of topological and metric," in *Robotics and Autonomous Systems*, 44(1), pp. 3-14, 2003.

[5] P. G. Zavlangas and S. G. Tzafestas, "Integration of topological and metric maps for indoor mobile robot path planning and navigation," *Proc. of the 2nd Hellenic Conference on AI: Methods and Applications of Artificial Intelligence*, pp. 121-130, 2002.

[6] J. Wolf, W. Burgard, H. Burkhardt, "Robust vision-based localization by combining an image retrieval system with Monte Carlo localization," *IEEE Transactions on Robotics*, 21(2), pp. 208-216, 2005.

[7] K. L. Ho and P. Newman, "Detecting loop closure with scene sequences," *Int. Journal of Computer Vision*, 74(3), pp. 261-286, 2007.

[8] S. Rady, A. Wagner, and E. Badreddin, "Entropy-based features for robust place recognition," *Proc. of IEEE Int. Conf. on System, Man and Cybernetics*, pp. 713-718, 2008.

[9] S. Rady, A. Wagner, and E. Badreddin, "Efficient codebook generation for appearance-based recognition," *Proc. of IEEE Int. Asian Control Conference*, pp. 1656-1661, 2009.

[10] E. Menegatti, T. Maeda, and H. Ishiguro, "Image-based memory for robot navigation using properties of the omnidirectional images," in *Robotics and Autonomous Systems*, 47(4), pp. 251-267, 2004.

[11] A. C. Murillo, J. J. Guerrero, and C. Sagues, "Surf features for efficient robot localization with omnidirectional images," *Proc. of IEEE Int. Conf. on Robotics and Automation*, pp. 3901-3907, 2007.

[12] J. Luo, A. Pronobis, B. Caputo, and P. Jensfelt, "Incremental learning for place recognition in dynamic environments," *Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 721-728, 2007.

[13] A. Tapus and R. Siegwart, "Incremental robot mapping with fingerprints of places," *Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 2429-2434, 2005.

[14] O. Martinez Mozos *et al.*, "Semantic labeling of places using information extracted from laser and vision sensor data," in *IEEE/RSJ IROS Workshop: From sensors to human spatial concepts,* 2006.

[15] I. Ulrich and I. Nourbakhsh, "Appearance-based place recognition for topological localization," *Proc. of IEEE Int. Conf. on Robotics and Automation*, pp. 1023-1029, 2000.

[16] P. Blaer and P. K. Allen, "Topological mobile robot localization using fast vision techniques," *Proc. of IEEE Int. Conf. on Robotics and Automation*, pp. 1031-1036, 2002.

[17] M. Artac, "Mobile robot localisation with incremental PCA", *Proc. of the 11th Mediterranean Electrotechnical Conference (MELECON),* pp. 192-197, 2002.

[18] M. M. Ullah *et al.*, "Towards robust place recognition for robot localization," *proc. of IEEE Int. Conf. on Robotics and Automation,* pp. 530-537, 2008.

[19] D. G. Lowe, "Distinctive image features from scale invariant keypoints," *Int. Journal of Computer Vision*, 2(60), pp. 91-110, 2004.

[20] F. Li and J. Kosecka, "Probabilistic location recognition using reduced feature set," *Proc. of IEEE Int. Conf. on Robotics and Automation*, pp. 3405-3410, 2006.

[21] J. Wang, H. Zha and R. Cipolla, "Coarse-to-fine vision-based localization by indexing scale-Invariant features," *IEEE Transactions of Systems, Man, and Cybernetics,* pp. 413-422, 2006.

[22] V. Chandola, A. Banerjee, and V. Kumar, "Outlier Detection-A Survey", *Tech. Report 07-017*, University of Minnesota, Aug. 2007.

[23] F. Jurie and B. Triggs, "Creating efficient codebooks for visual recognition," *Proc. of IEEE Int. Conf. on Computer Vision*, pp. 604-610, 2005.

[24] "The COLD (COsy Localization Database) database." [Online]. Available: http://cogvig.nada.kth.se/COLD/