

# Towards Robust Place Recognition for Robot Localization

M. M. Ullah<sup>\*†</sup>, A. Pronobis<sup>‡</sup>, B. Caputo<sup>\*†</sup>, J. Luo<sup>\*†</sup>, P. Jensfelt<sup>‡</sup>, and H. I. Christensen<sup>‡§</sup>

<sup>\*</sup>IDIAP Research Institute  
1920 Martigny, Switzerland

<sup>‡</sup>Centre for Autonomous Systems  
Royal Institute of Technology

<sup>§</sup>College of Computing  
Georgia Institute of Technology  
Atlanta, GA 30332-0760, USA  
hic@cc.gatech.edu

<sup>†</sup>EPFL, 1015 Lausanne, Switzerland

SE-100 44 Stockholm, Sweden

[mullah,bcaputo,jluo]@idiap.ch

[pronobis, patric]@kth.se

**Abstract**—Localization and context interpretation are two key competences for mobile robot systems. Visual place recognition, as opposed to purely geometrical models, holds promise of higher flexibility and association of semantics to the model. Ideally, a place recognition algorithm should be robust to dynamic changes and it should perform consistently when recognizing a room (for instance a corridor) in different geographical locations. Also, it should be able to categorize places, a crucial capability for transfer of knowledge and continuous learning. In order to test the suitability of visual recognition algorithms for these tasks, this paper presents a new database, acquired in three different labs across Europe. It contains image sequences of several rooms under dynamic changes, acquired at the same time with a perspective and omnidirectional camera, mounted on a socket. We assess this new database with an appearance-based algorithm that combines local features with support vector machines through an ad-hoc kernel. Results show the effectiveness of the approach and the value of the database.

## I. INTRODUCTION

A valuable competence for a robot is to know its position in the world, i.e. the ability to localize. This topic is vastly researched, with methods spanning from geometrical [1], to topological [2], and hybrid [3]. While traditionally sonar and/or laser have been the sensory modalities of choice [4], recent advances in vision have made this option more interesting, as it provides richer information for loop closing, recovery from the kidnapped robot problem and a way to introduce contextual information into the system.

From the point of view of place recognition, there are several open challenges for using vision robustly, that can be summarized as follows:

- 1) *robustness to dynamic changes*; The visual appearance of places varies in time because of illumination changes (day and night, artificial light on and off) and because of furniture moved around, objects being taken out of drawers, and so on. We call these changes dynamic because they are visible just when considering a room across a span of time of at least several hours.

This work was supported by the EU integrated projects CoSy FP6-004250-IP, www.cognitivesystems.org (MMU, AP, PJ) and DIRAC IST-027787, www.diracproject.org (BC, JL), and the Swedish Research Council contract 2005-3600-Complex (AP). The support is gratefully acknowledged. Very special thanks go to the colleagues that made it possible the acquisition of the COLD database: Óscar Martínez Mozos and Wolfram Burgard in Freiburg, Matej Artač, Aleš Štítec and Aleš Leonardis in Ljubljana as well as Hendrik Zender and Geert-Jan Kruijff in Saarbrücken.

A visual place classification algorithm should be able to tackle effectively these variations.

- 2) *robustness to geographical changes*; The same room class (such as ‘corridor’, ‘bathroom’, etc) will look different in different geographic locations, while still preserving some distinctive common visual features. We call these changes *geographical* because they are visible only by physically locating the robot platform in two different environments. Ideally, when training and testing on data acquired in lab ‘A’ one should get a performance consistent with that achieved by training and testing on data acquired in lab ‘B’.
- 3) *robustness to categorical changes*; Humans are able to recognize a room as ‘an office’, or ‘a kitchen’ or ‘a corridor’, even if they see it for the first time. This is because they are able to build categorical models of places. A visual place recognition algorithm should be able to categorize by building models on features that are distinctive of the category at hand, across various instances of places. This information would be extremely valuable for knowledge transfer regarding places and their functionalities, thus adding rich contextual information.

A major obstacle to research these issues is the difficulty to test algorithms with respect to these challenges. While it would be possible to move robots from one place to another for a long span of time, this would not allow fair comparison between methods. This paper aims at filling this gap, and presents a new database that we call COsy Localization Database (COLD). The COLD database was acquired in three different labs across Europe, imaging several rooms under dynamic changes. For the acquisition procedure we used perspective and omnidirectional cameras, mounted together on a socket, that was moved from one lab to another. The socket was mounted on the robot platform available at each lab, and each robot was driven manually across several rooms for the data acquisition. To the best of our knowledge, the COLD database is the biggest and most varied database for robot localization in indoor settings. The database is freely available to the community via the Internet [5].

We assessed the database using a purely appearance-based method that proved successful for indoor place recognition

[6], [7], [8]. The algorithm uses local descriptors to extract rich visual information from image sequences, and support vector machines [9] for the classification step.

The rest of the paper is organized as follows: after a review of related literature (Section II), we introduce our new database (Section III). Section IV describes our recognition algorithm and Section V reports the thorough experiments showing the value of the proposed approach. We draw conclusions and discuss future research in Section VI.

## II. RELATED WORK

Topological localization is a vastly researched topic in the robotic community [4], [10], [11], [12], [13], where vision and laser range sensors are usually the privileged modalities. Vision-based approaches employ either perspective [14], [15] or omnidirectional cameras [16], [17], [18]. They can be roughly divided into landmark-based approaches, where localization is based on artificial or natural landmarks [19], [20], [18], [15], and methods employing global image features [16], [10], [11], [14], [17], [8]. Robustness to dynamic changes has not been investigated much so far, with the notable exceptions of [7], [6]. The same applies for the categorization problem, where first attempts based on laser range cues and visual sensors showed promising results [4] for the semantic labeling of places into three categories (corridor, room, doorway). The problem was also addressed by Torralba *et al.* [14] who studied the issue of categorizing novel places based on global image features (“gist”). Still the place categorization problem is far from being solved. We are not aware of previous work explicitly addressing the issue of robustness with respect to geographical changes.

The main contribution of this paper is the creation and assessment of an extensive database for robot localization, that can be used for benchmark evaluation. There are a number of heavily used databases in robotics [21], [22] and computer vision [23], [24], [25]. In robotics, these databases are used mainly for testing algorithms for simultaneous localization and mapping (SLAM) and mostly contain odometry and range sensor data. A notable exception is the recently introduced IDOL2 database [26], that can be seen as a preliminary attempt to provide the community with a database for visual place recognition under dynamic changes. Compared to the COLD database, the IDOL2 provides image sequences only from a perspective camera, and images 5 rooms in the same laboratory across a time span of roughly 6 months, under varying illumination conditions. The IDOL2 database can thus be regarded as especially devoted to the study of robustness under dynamic changes. The database presented in this paper makes an important contribution by providing data from vision (perspective and omnidirectional cameras) and range sensors (laser scanner). In addition, the data is labeled with the position at which it was acquired which makes it ideal for benchmarking place recognition algorithms. The introduction of standard benchmark databases has made an impact on the research on the SLAM problem, allowing different methods to be more fairly compared in the same scenario. The authors hope that similarly COLD

will become a standard dataset and will boost the research on place recognition and localization.

## III. THE COLD DATABASE

The COLD (COsy Localization Database) database is a new collection of image sequences. It represents an effort to provide a flexible testing environment for evaluating vision-based place recognition systems aiming to work on mobile platforms in real-world environments. The COLD database consists of three separate sub-datasets, acquired at three different indoor labs, located in three different European cities: the Visual Cognitive Systems Laboratory at the University of Ljubljana, Slovenia; the Autonomous Intelligent System Laboratory at the University of Freiburg, Germany; and the Language Technology Laboratory at the German Research Center for Artificial Intelligence in Saarbrücken, Germany. For each lab, we acquired image sequences of several rooms. We always used the same camera settings, consisting of a perspective and omnidirectional cameras, mounted together on a portable socket (as shown in Fig. 3d). The socket with the two cameras was moved from one lab to another, and mounted on the mobile platform available at each place. Sequences were acquired under different weather and illumination conditions, and across a time span of two/three days. Special care was put in the choice of the rooms to image, and for each lab there exists a set of sequences containing rooms with similar functionalities that are also contained in the other two. Thus, the COLD database is an ideal testbed for assessing the robustness of visual place recognition algorithms with respect to dynamic, geographical and categorical changes. We are not aware of other databases available to the robotic research community that contain image sequences of indoor environments, acquired under dynamic changes and in different laboratories. The database is available through the web and can be downloaded from <http://cogvis.nada.kth.se/COLD>. The database is currently being expanded, and similar image sequences are going to be acquired at the Computational Vision and Active Perception Laboratory at the Royal Institute of Technology in Stockholm, Sweden. Due to time constraints, we did not assess the omnidirectional sequences yet, and therefore we will not report experimental results on them in the paper. From now onwards, we will refer to the three sub-databases, taken by the perspective camera, with the names of the cities where the labs were imaged (COLD-Saarbrücken, COLD-Freiburg and COLD-Ljubljana).

In the rest of the section we describe the acquisition setup, which was specific to each of the three locations (Section III-A). Then, Section III-B explains the acquisition procedure that we followed. And finally, Section III-C summarizes our annotation methods. For further details, we refer the reader to [5].

### A. The Acquisition Setup

For the image sequence acquisition, we tried to select rooms that are common to most of modern lab environments, for instance the kitchen, the printer area and the corridor.

Laboratory	Corridor	Terminal room	Robotics lab	1-person office	2-persons office	Conference room	Printer area	Kitchen	Bath room	Large office	Stairs area	Lab
Saarbrücken	XO△□	△	△	O△□	X△	△	XO△□	□	XO△□			
Freiburg	XO△			O△	XO△		X△	△	XO△	△	XO△	
Ljubljana	X△				X△		X△	△	X△			X△

TABLE I A list of different types of rooms that were imaged at the three labs. Each room is marked with different shapes according to the sequences in which it was included: ‘X’ stands for standard sequence A; ‘O’ stands for standard sequence B; ‘△’ stands for extended sequence A; and ‘□’ stands for extended sequence B. See Section III-B for more details on the sequence acquisition and the naming convention adopted here.

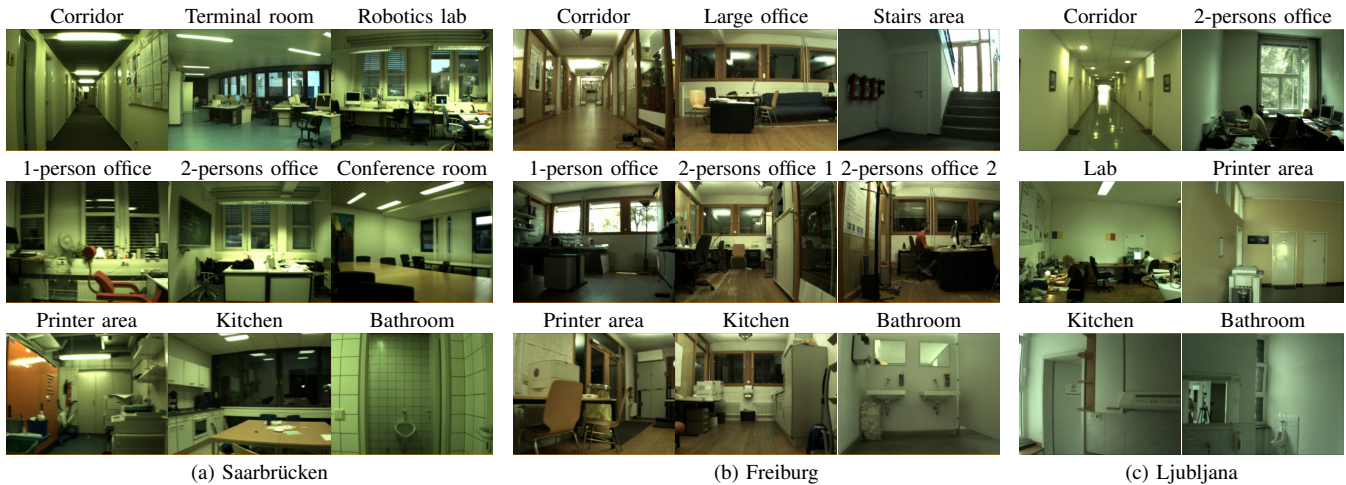


Fig. 1. Examples of images of the three labs acquired by the perspective camera showing the interiors of the rooms.

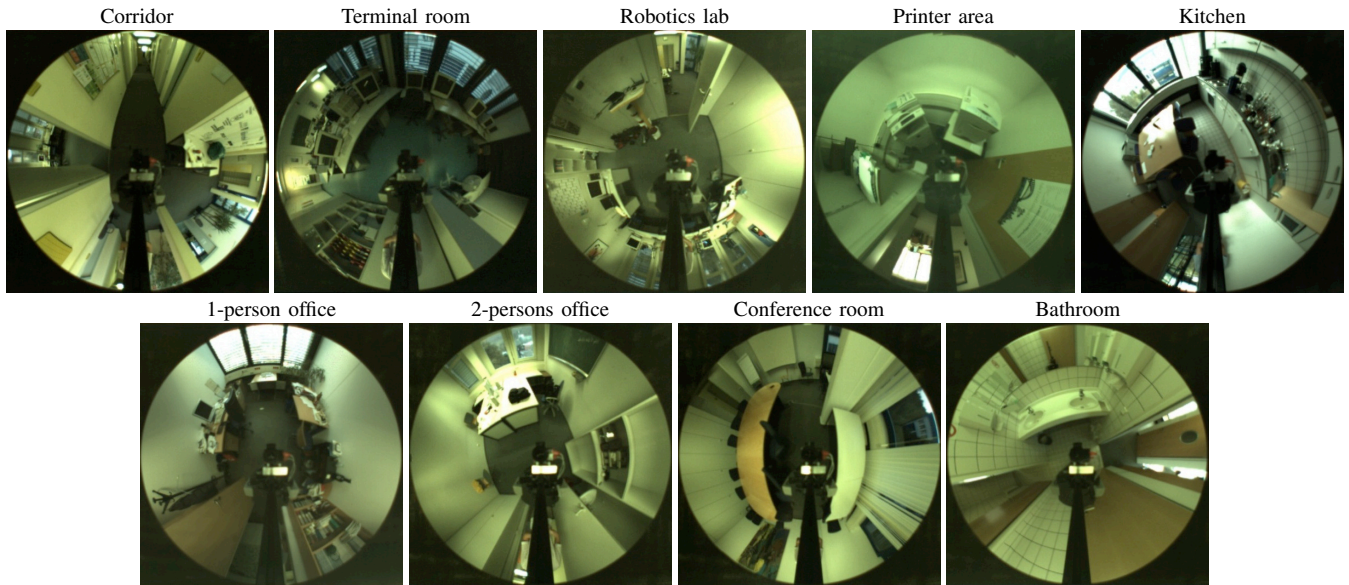


Fig. 2. Examples of images of Saarbrücken acquired by the omnidirectional camera showing the interiors of the rooms.

However, some rooms were specific to particular labs, like the terminal room and the robotics lab room in Saarbrücken. Table I provides a list of rooms that were imaged at the three labs, as well as the types of sequences these rooms correspond to. Sample images of each room taken by the perspective and omnidirectional camera, for each lab, are shown in Fig. 1-2. From the Freiburg images we can see that the separating walls between offices and rooms are made of glass; this will likely make these sequences challenging. Saarbrücken and Ljubljana instead have concrete walls.

At each lab, a different mobile platform, equipped with the very same cameras, was employed for image acquisition. The camera setup is built using two Videre Design MDCS2 digital cameras, one for perspective images, the other for the omnidirectional images. The catadioptric omnidirectional vision system was constructed using a hyperbolic mirror. The heights of the cameras varied on each of the three mobile platforms, because of the differences between the robots. Fig. 3d presents the three mobile platforms employed during image acquisition. All the images were acquired with

LAB	Standard sequences						Extended sequences					
	Cloudy		Night		Sunny		Cloudy		Night		Sunny	
	A	B	A	B	A	B	A	B	A	B	A	B
Saarb.	3	5	3	3	-	3	3	3	3	-	3	
Freib.	3	3	3	-	4	3	3	-	3	-	4	-
Ljubl.	3	-	3	-	3	-	3	-	3	-	3	-

TABLE II Acquisition results for each of the three laboratories. Two different portions of the laboratories are annotated as ‘A’ and ‘B’.

a resolution of  $640 \times 480$  pixels, with the auto-exposure and the auto-focus modes turned on. The lens of the perspective camera had a wider view angle ( $84.9^\circ \times 68.9^\circ$ ) than the lens of the omnidirectional camera ( $56.1^\circ \times 43.6^\circ$ ).

### B. The Acquisition Procedure

We followed the same procedure during image acquisition at each lab. The robot was manually driven (at a speed of roughly 0.3m/s) through each of the available rooms while continuously acquiring images at the rate of 5 frames per second. Since the two cameras were synchronized, for every perspective image, there is an omnidirectional image with the same time stamp. For the different weather and illumination conditions (cloudy, night and sunny), the acquisition procedure was repeated at least thrice, resulting in a minimum of three image sequences, acquired one after the other, under the same illumination condition.

At each lab, different paths were followed by the robot during image acquisition: (a) the *standard* path, in case of which the robots were driven across rooms that are most likely to be found in most labs; (b) the *extended* path, in case of which the robots were additionally driven across the rooms that were specific for each lab. For Saarbrücken and Freiburg, there are two portions of the lab, which were treated separately. As a result, two different sets of sequences (annotated as A and B) were acquired. Detailed information about the number of sequences in the database for each lab, portion and illumination setting can be found in Table II. Due to manual control of the robot, differences in viewpoints still can be seen between different sequences, even if they come from the same acquisition path. Fig. 3a-c presents the two types of paths that the robot followed at each lab (for Saarbrücken and Freiburg, only one portion is presented). The total number of frames in each image sequence depends on the lab and the path that the robot followed (roughly 1000-2800 for Saarbrücken, 1600-2800 for Freiburg and 2000-2700 for Ljubljana).

### C. The Data Annotation

For labeling the images, we followed the same procedure as [26]: the pose of the robot was estimated during the acquisition process using a laser-based localization technique. Each image was then labeled with the exact pose of the robot at the moment of acquisition and assigned to one of the available rooms according to the position. This strategy could not be followed in Ljubljana, because the available robot platform did not have a laser scanner. Thus, for the Ljubljana sequences, the annotation process was done using

the odometry data with manual corrections. The laser scans and odometry data are provided together with the database.

For the perspective camera, an important consequence of this annotation procedure is that the label assigned to a frame might be weakly related to its visual content due to the constrained field of view. This is particularly true for the Freiburg sequences, because the walls in that laboratory are mostly made of glass. We can thus expect that the Freiburg sub-database will be particularly challenging. However, we believe that this challenge might be tackled by using the omnidirectional camera.

## IV. ROBUST PLACE RECOGNITION

A robust visual place recognition algorithm needs to combine descriptive, discriminative and generalization abilities. In order to capture these properties, we used a fully supervised, appearance-based approach that has shown good performance on the place recognition problem [7], [6], [8]. Each room was represented during training by a collection of frames capturing its visual appearance under varying viewpoints, and possibly under varying acquisition conditions. Local features were extracted from the training images using a Harris-Laplace detector [27] and the SIFT descriptor [28]. These features provide an excellent trade-off between descriptive power, thanks to the local SIFT descriptors, and generalization abilities, as their local nature makes them capture significant fragments which are likely to appear again in different settings. For the classification step we used support vector machines (SVMs, [9]). As SVMs require the computation of scalar products on the feature vectors, special care must be used in choosing an appropriate kernel function. Here we used the match kernel [29], that has shown good performance in several visual recognition domains. Given two local features  $L_h$  and  $L_k$ , the match kernel is defined as

$$K(L_h, L_k) = \frac{1}{n_h} \sum_{j_h=1}^{n_h} \max_{j_k=1, \dots, n_k} \left\{ K_l(L_h^{j_h}, L_k^{j_k}) \right\}, \quad (1)$$

where the local feature similarity kernel  $K_l$  consists of any Mercer kernel, acting on the local SIFT descriptors. Once the algorithm is trained, it can be used to recognize images from sequences as belonging to one of the places seen during the training stage. The goal is to recognize correctly each single image seen by the system and the recognition is based on one image only. As it will be shown in the next section, this algorithm is able to perform robust place recognition with respect to dynamic, geographical and categorical changes.

## V. EXPERIMENTS

We assessed the COLD database with two series of experiments. In the first series, we ran three sets of experiments, one for each lab. For each set, training and testing was always done on different sequences acquired in the same lab. We trained on one illumination condition, and tested on sequences acquired under various illumination conditions, and after some time. With these experiments we were able to address at the same time the robustness with respect to dynamic and geographical changes; results are reported in

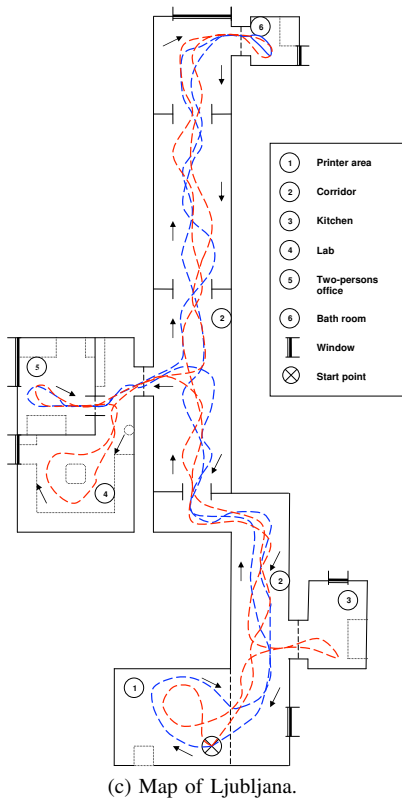
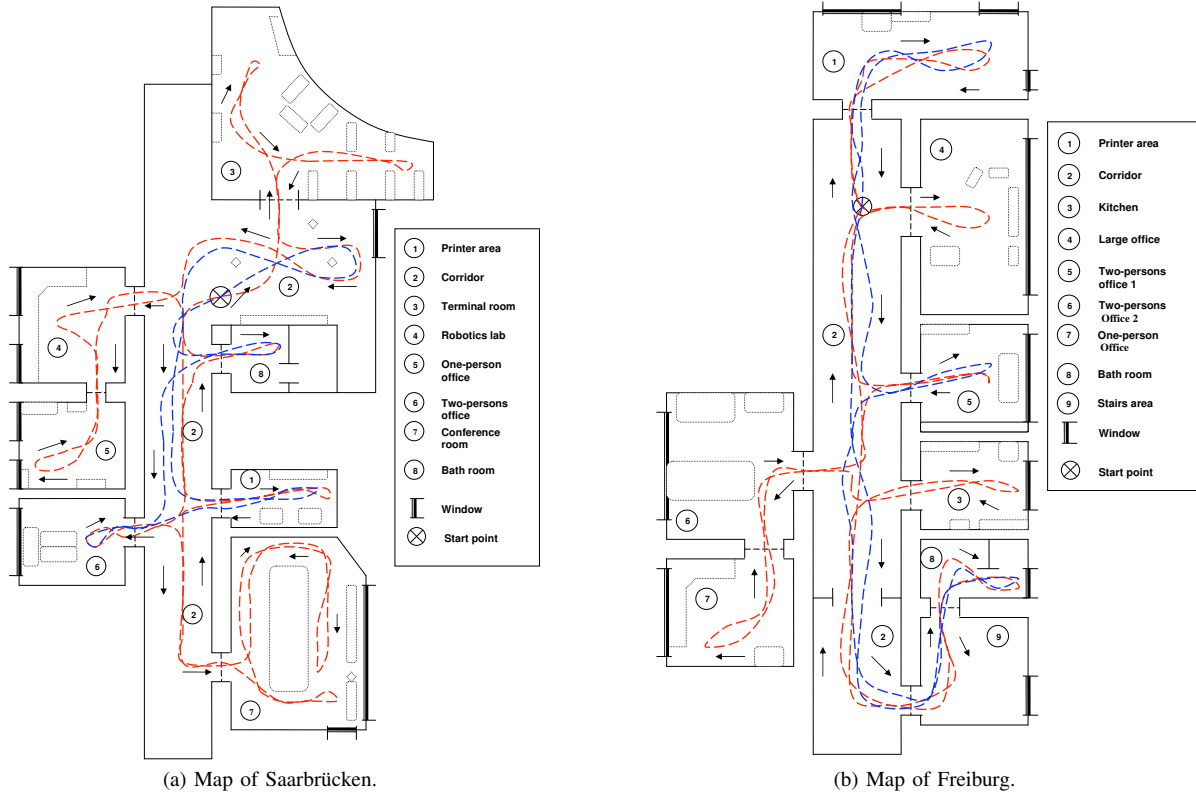


Fig. 3. Two different types of paths followed by the robot at each lab. The standard path is represented with blue dashes and the extended path is represented with red dashes, for each of the three labs (Fig. 3a-c). Arrows indicate the direction of driving of the robot. The three different mobile platforms employed for image acquisition at the three labs, are shown in Fig. 3d.

Section V-A. We then addressed the robustness to categorical changes in the second series of experiments. We chose image sequences containing the same rooms for each lab, and we trained on sequences from two labs and tested on the remaining one. Experiments were repeated on increasingly challenging data, because of dynamic changes, and for all possible permutations of training and test sets. These results are reported in Section V-B. For all the experiments, we used our extended version of the *libsvm* [30] library, and we determined the SVM and kernel parameters via cross-validation. The cross-validation was performed on the COLD-Saarbrücken sequences, and the obtained values of parameters were used for all the experiments. The experiments were conducted several times using all the possible permutations of the training and test sequences; average results are reported with standard deviations.

#### A. Experiments on Dynamic and Geographical Changes

The task during these experiments was to recognize a room, seen during training, when imaged under different conditions, i.e. at a different time and/or under different illumination settings. We performed experiments for sequences from each laboratory. For each experiment, training set consisted of one sequence taken in one laboratory, and testing was done on sequences acquired in the same laboratory, under various conditions. With these experiments it was possible to verify robustness to dynamic changes, due to the selection of the training and test sets, as well as to geographic changes, as the parameters of the algorithms were always the same.

The obtained results for all the three labs are presented in Fig. 4a (COLD-Saarbrücken), Fig. 4b (COLD-Freiburg) and Fig. 4c (COLD-Ljubljana). For each lab, the bar chart in the top row reports results for the standard sequences; results reported in the bottom row are for the extended sequences. For each training illumination condition (indicated on top of the charts), the bars present the average classification rates over the corresponding testing sequences under the illumination condition marked on the bottom axis. We will first comment on the results from the point of view of dynamic changes. Then we will study the robustness to geographic changes.

*a) Robustness to Dynamic Changes:* It can be observed that the method achieves very good performance when trained and tested under stable illumination conditions. On average, the system classified correctly 90.5% of images from standard sequences and 83.8% of images from extended sequences acquired in Saarbrücken, 85.6% and 81.8% of images from sequences acquired in Freiburg, and 90.4% and 85.5% of images from sequences acquired in Ljubljana. Note that even if the illumination conditions for training and testing were the same, the algorithm had to tackle other kinds of variability introduced e.g. by human activity or viewpoint changes due to the manual control of the robot. The errors usually occur in the transition areas between the rooms. We could also observe that the classification rates obtained for the standard image sequences are generally better than those

obtained for the extended sequences. This can be explained by the fact that the extended sequences contain a larger number of classes (rooms), which makes the problem harder.

It can be seen from Fig. 4 that the system achieves good performance also when testing is performed on sequences acquired under different illumination conditions than those used for training. In general, the best recognition rates were obtained for training sequences acquired during cloudy weather. Consider for example the COLD-Ljubljana results for the standard path (Fig. 4c, top). For this experiment, the average classification rate was equal to 83.46% for night test sequences and 84.01% for sunny test sequences. This is close to the 85.88% achieved for experiments in case of which training and testing was done under similar illumination.

*b) Robustness to Geographic Changes:* The baseline method provides a good robustness to geographic changes. When considering all the results obtained by training and testing on similar illumination conditions, we get an average classification rate of 87.5% for COLD-Saarbrücken, 83.70% for COLD-Freiburg and 87.95% for COLD-Ljubljana. These results are very consistent. At the same time, we can observe that there is a decrease in performance for COLD-Freiburg. This can be caused by the glass walls in Freiburg and the fact that the cameras were mounted significantly lower than in case of the other two labs, resulting in less diagnostic information in some of the images. A similar behavior, but for lower overall performance, can be observed in case of the experiments under varying illumination conditions. Here we achieve classification rates of 75.82% for COLD-Saarbrücken, 71.03% for COLD-Freiburg and 79.98% for COLD-Ljubljana. We underline once again the lower performance on the COLD-Freiburg data, which confirms that this collection is the most challenging of the whole COLD database.

#### B. Experimental Results on Categorical Changes

For these experiments, our motivation was to explore whether our baseline method is able to build categorical models of places. The underlying assumption here is that rooms within the same category will share a certain degree of visual similarity, because of functional objects and their layout (for instance, the printer machines in printer areas). Fig. 5 shows some frames from the sequence part ‘office’ and ‘printer’, from the three different labs. It is very interesting to note that, even if the furniture style differs from lab to lab, there still is a clear visual similarity between these frames, that should make it possible for our appearance-based baseline method to perform reasonably well. Another interesting point is that those visually similar frames have been captured, for each sequence from each lab, in very different geometric positions (Fig. 5).

In order to perform these experiments, we selected four different rooms (corridor, printer area, two-persons office and bathroom), all available in the standard sequences of each laboratory. The algorithm was trained on two sequences taken from two laboratories, and it was tested on sequences taken at the third remaining laboratory. This time the algo-

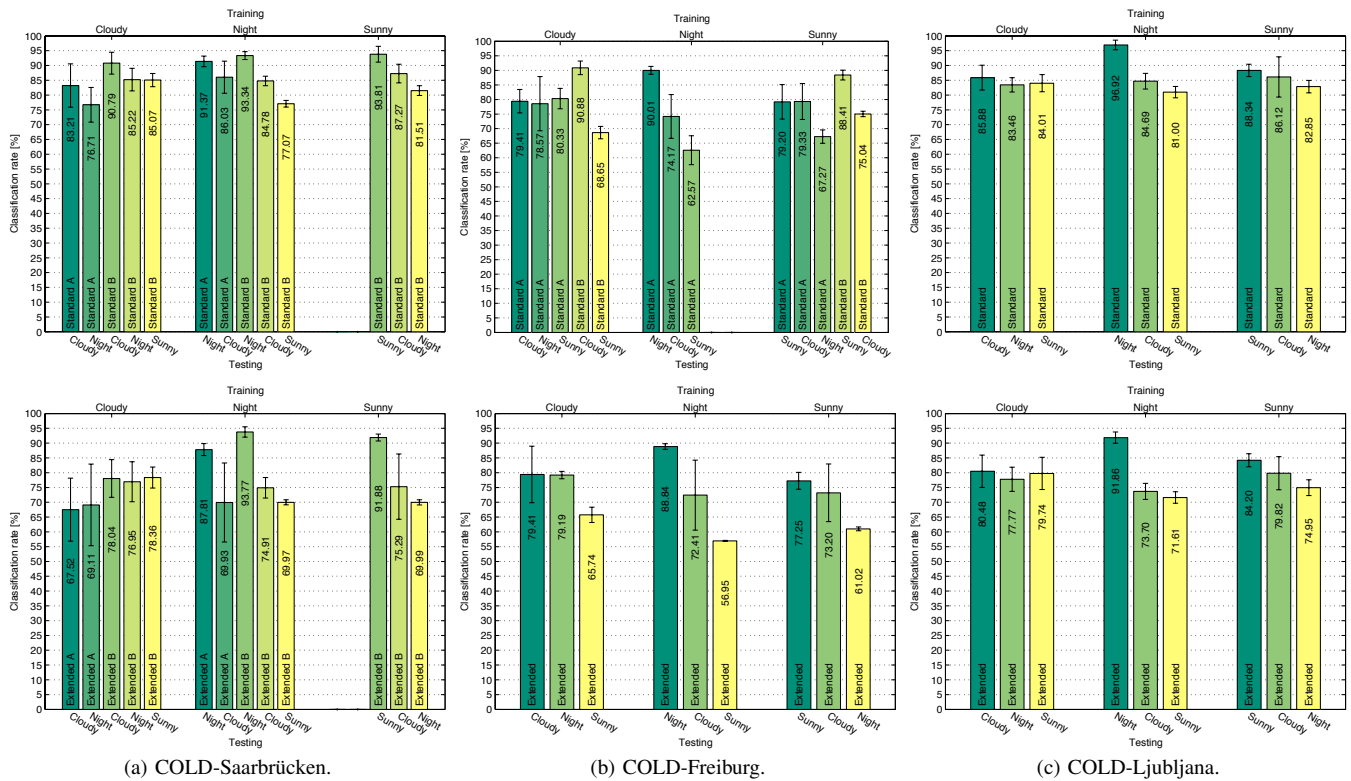


Fig. 4. Average results of the experiments with the three sub-databases. The results for the standard sequences (top row) and extended sequences (bottom row) are given separately for each sub-database. The classification rates are grouped according to the type of illumination conditions under which the training sequences were acquired. The bottom axes indicate the illumination conditions used for testing. The uncertainties are given as one standard deviation. Results corresponding to the two different portions of the labs are indicated by 'A' and 'B'.

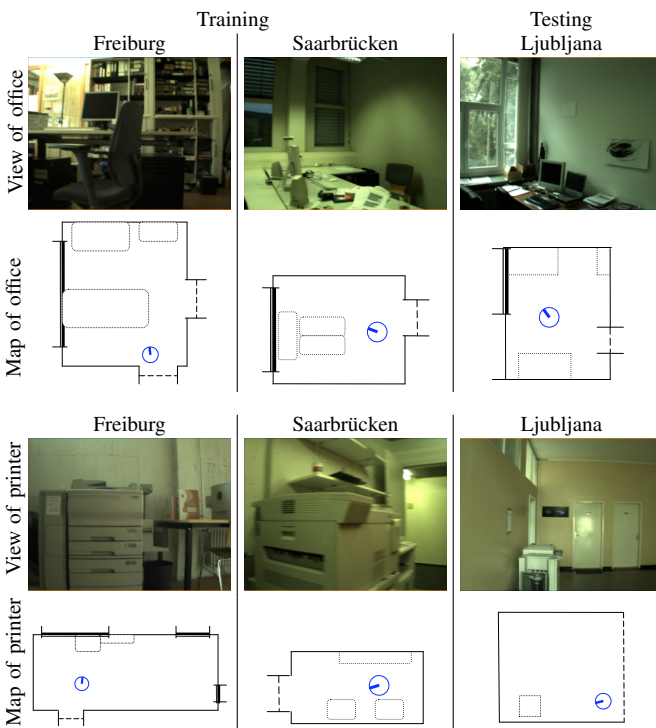


Fig. 5. Examples of images of rooms within the same category from the three different labs. The acquisition poses were marked with 'o' on the maps below the images.

gorithm was not optimized for the SVM and kernel parameters. Each room was treated separately during the testing phase. To eliminate other unnecessary influences, we took, for training and test, sequences with the same illumination conditions. Fig. 6 shows the average classification rates of the experiments under two weather conditions (cloudy and night). On average, the system achieved quite good classification rate for the corridor (76.37%), compared to any other room. A possible reason could be the fact that the visual appearance of the corridor is very similar in the two labs (Saarb. and Ljubl.). Given that we did not perform any parameter optimization for these experiments and that quite many frames in the sequences were not categorical informative, the results are quite acceptable.

### C. Discussion

The baseline results of the extensive experimental evaluation presented here show that the COLD database is well suited for testing visual recognition algorithms for robust robot localization. While the experimental procedures we proposed is of course only one of the many possible ways to use these data, they address effectively some important properties that are desirable for visual localization. The results obtained point out quite clearly that categorization is still an unsolved problem. The experimental results reported in this section will be made available on the database web-page to facilitate benchmarking between methods and will make the

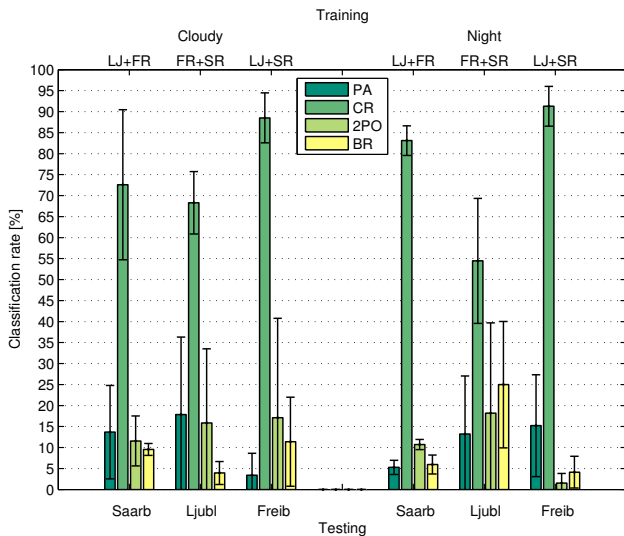


Fig. 6. Average results of the categorization experiments for each room, under two weather conditions. Training is performed on the sequences from two sub-databases whereas testing on the sequence from the remaining third sub-database. The sub-databases are marked as: ‘SR’ for Saarbrücken, ‘FR’ for Freiburg, and ‘LJ’ for Ljubljana.

COLD database a useful resource to the community.

## VI. CONCLUSIONS AND FUTURE WORKS

This paper addressed the issue of robust visual place recognition for robot localization. We considered robustness with respect to dynamic, geographical and categorical changes. We presented a database, called CoSy Localization Database (COLD), consisting of image sequences acquired under varying illumination conditions and across a time span of several days. Sequences have been acquired in three different labs across Europe using perspective and omnidirectional cameras, mounted together on a socket. We assessed the database with a comprehensive series of experiments, using an appearance-based visual recognition approach that proved successful for the localization task.

This work can be extended in many ways. From the point of view of the database, we are acquiring a fourth set of sequences at the Computational Vision and Active Perception laboratory in Stockholm. The place recognition algorithm could also be improved in several directions. The robustness of the method for both recognition and categorization can be increased by incorporating knowledge from several visual cues or even different modalities (here both visual sensors and laser range scanner) as it was done in [6], [31]. Additionally, information about the confidence of the final decision could be utilized to further improve reliability [6]. Also, the possibility to categorize should be exploited for knowledge transfer, as proposed in [32], and for life long learning, as suggested in [7], [33]; we intend to explore both possibilities.

## REFERENCES

[1] M. Jogan and A. Leonardis, “Robust localization using an omnidirectional appearance-based subspace model of environment,” *Robotics and Autonomous Systems*, vol. 45, no. 1, October 2003.

[2] I. Ulrich and I. R. Nourbakhsh, “Appearance-based place recognition for topological localization,” in *Proc. ICRA’00*.

[3] S. Thrun, “Learning metric-topological maps for indoor mobile robot navigation,” *Artificial Intelligence*, vol. 1999, no. 1, 1998.

[4] O. Martínez Mozos, C. Stachniss, and W. Burgard, “Supervised learning of places from range data using adaboost,” in *Proc. ICRA’05*.

[5] “The COLD (CoSy Localization Database) database.” [Online]. Available: <http://cogvis.nada.kth.se/COLD/>

[6] A. Pronobis and B. Caputo, “Confidence-based cue integration for visual place recognition,” in *Proc. IROS’07*.

[7] J. Luo, A. Pronobis, B. Caputo, and P. Jensfelt, “Incremental learning for place recognition in dynamic environments,” in *Proc. IROS’07*.

[8] A. Pronobis, B. Caputo, P. Jensfelt, and H. I. Christensen, “A discriminative approach to robust visual place recognition,” in *Proc. IROS’06*.

[9] N. Cristianini and J. S. Taylor, *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press, 2000.

[10] J. Gaspar, N. Winters, and J. Santos-Victor, “Vision-based navigation and environmental representations with an omni-directional camera,” *IEEE Trans RA*, vol. 16, no. 6, 2000.

[11] P. Blaer and P. Allen, “Topological mobile robot localization using fast vision techniques,” in *Proc. ICRA’02*.

[12] H. Andreasson, A. Treptow, and T. Duckett, “Localization for mobile robots using panoramic vision, local features and particle filter,” in *Proc. ICRA’05*.

[13] B. Kuipers and P. Beeson, “Bootstrap learning for place recognition,” in *Proc. AAAI’02*.

[14] A. Torralba, K. P. Murphy, W. T. Freeman, and M. A. Rubin, “Context-based vision system for place and object recognition,” in *Proc. ICCV’03*.

[15] H. Tamimi and A. Zell, “Vision based localization of mobile robots using kernel approaches,” in *Proc. IROS’04*.

[16] I. Ulrich and I. R. Nourbakhsh, “Appearance based place recognition for topological localization,” in *Proc. ICRA’00*.

[17] E. Menegatti, M. Zoccarato, E. Pagello, and H. Ishiguro, “Image-based monte-carlo localisation with omnidirectional images,” *Robotics and Autonomous Systems*, vol. 48, no. 1, 2004.

[18] A. C. Murillo, J. J. Guerrero, and C. Sagues, “Surf features for efficient robot localization with omnidirectional images,” in *Proc. ICRA’07*.

[19] M. Mata, J. M. Armingol, A. de la Escalera, and S. M. A., “Using learned visual landmarks for intelligent topological navigation of mobile robots,” in *Proc. ICRA’03*.

[20] S. Se, D. Lowe, and J. Little, “Vision-based mobile robot localization and mapping using scale-invariant features,” in *Proc. ICRA’01*.

[21] A. Howard and N. Roy, “The Robotics Data Set Repository (Radish),” 2003. [Online]. Available: <http://radish.sourceforge.net/>

[22] E. Nebot, “The Sydney Victoria Park dataset.” [Online]. Available: <http://www-personal.acfr.usyd.edu.au/nebot/dataset.htm>

[23] G. Griffin, A. Holub, and P. Perona, “Caltech-256 Object Category Dataset,” Caltech, Tech. Rep. 7694, 2007. [Online]. Available: <http://authors.library.caltech.edu/7694/>

[24] “The PASCAL Visual Object Classes challenge.” [Online]. Available: <http://www.pascal-network.org/challenges/VOC/>

[25] “The KTH-TIPS image database.” [Online]. Available: <http://www.nada.kth.se/cvap/databases/kth-tips/>

[26] J. Luo, A. Pronobis, B. Caputo, and P. Jensfelt, “The IDOL2 database,” KTH, CAS/CVAP, Tech. Rep. 304, 2006. [Online]. Available: <http://cogvis.nada.kth.se/IDOL2/>

[27] C. Harris and M. Stephens, “A combined corner and edge detector,” in *Proc. AVS88*.

[28] D. G. Lowe, “Distinctive image features from scale-invariant key-points,” *IJCV*, vol. 60, no. 2, 2004.

[29] C. Wallraven, B. Caputo, and A. Graf, “Recognition with local features: the kernel recipe,” in *Proc. of ICCV’03*.

[30] C. C. Chang and C. J. Lin, *LIBSVM: A Library for Support Vector Machines*, 2001. [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>

[31] A. Pronobis, O. Martínez Mozos, and B. Caputo, “SVM-based discriminative accumulation scheme for place recognition,” in *Proceedings of ICRA’08*.

[32] J. Luo, A. Pronobis, and B. Caputo, “Svm-based transfer of visual knowledge across robotic platforms,” in *Proc. ICVS’07*.

[33] F. Orabona, C. Castellini, B. Caputo, J. Luo, and G. Sandini, “Indoor place recognition using online independent support vector machines,” in *Proc. BMVC’07*.