COLD: The COsy Localization Database

A. Pronobis*, B. Caputo†

* CAS/CVAP, Royal Institute of Technology
SE-100 44 Stockholm, Sweden
pronobis@csc.kth.se

† Idiap Research Institute
CH-1920 Martigny, Switzerland
bcaputo@idiap.ch

Abstract— **Two key competences for mobile robotic systems are localization and semantic context interpretation. Recently, vision became the modality of choice for these problems as it provides richer and more descriptive sensory input. At the same time, designing and testing vision-based algorithms still remains a challenge, as large amounts of carefully selected data are required to address the high variability of visual information. This paper presents a freely available database which provides a large-scale, flexible testing environment for vision-based topological localization and semantic knowledge extraction in robotic systems. The database contains 76 image sequences acquired in three different indoor environments across Europe. Acquisition was performed with the same perspective and omnidirectional camera setup, in rooms of different functionality and under various conditions. The database is an ideal testbed for evaluating algorithms in real-world scenarios with respect to both dynamic and categorical variations.**

I. Introduction

A major challenge to research on vision-based localization in mobile robotics is the difficulty to test the robustness of algorithms in presence of various visual variations. Since the results depend greatly on the input sensory data, which are inherently unstable over time, it is hard to measure the influence of the different parameters on the overall performance of the system. For the same reason, it is nearly impossible to compare fairly solutions which are usually evaluated in different environments, in different conditions, and under different assumptions. There is a need for standardized benchmarks and databases which would allow for such comparisons, simplify the experimental process and boost progress in the field.

Databases are heavily exploited in the computer vision community, especially for object recognition and categorization [1], [2], [3]. Also in robotics, research on Simultaneous Localization and Mapping (SLAM) makes use of several publicly available datasets [4], [5]. These are, however, primarily targeted for metric mapping problems and mostly contain odometry and range sensor data. A notable exception is the IDOL2 database [6], that can be seen as a preliminary attempt to create a database for visual place classification under dynamic changes.

This paper presents the COsy [7] Localization Database (COLD), a new collection of annotated data sequences acquired using visual and laser range sensors on a mobile platform. The database represents an effort to provide a large-scale, flexible testing environment for evaluating mainly vision-based topological localization and semantic knowledge extraction methods aiming to work on mobile robots in realistic scenarios. Thanks to our design choices, COLD is also a valuable source of data for metric mapping problems.

The COLD database consists of three separate sub-datasets, acquired at three different indoor laboratory environments, 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. The sequences in the database were recorded using several mobile robots and both perspective and omnidirectional cameras. Laser range scans and odometry data were also captured for most of the sequences. At each laboratory, data acquisition was performed within several rooms using the same camera setup. The cameras were mounted on a portable bracket which was moved from one laboratory to the other and attached to the mobile platform available at each place. Image sequences were acquired under different illumination conditions and across several days. Special care was taken 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, COLD is an ideal testbed for assessing the robustness of localization and recognition algorithms with respect to both dynamic and categorical changes. To the best of our knowledge, it is the largest and most comprehensive database for robot localization in indoor settings. From now onwards, we will refer to the three sub-datasets as COLD-Saarbrücken, COLD-Freiburg and COLD-Ljubljana.

The COLD database, annotation files, and tools for data processing are freely available via the Internet [8].

The rest of the paper is organized as follows: Section II discusses further our motivations for building the database, the design and possible application scenarios. Sections III and IV describe the acquisition setup, procedure and the ac-

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

quisition outcome. The annotation and data post-processing are described in Section V. We draw conclusions in Section VI. For further details on the database and a description of the baseline evaluation, we refer the reader to [8], [9].

II. Motivation and Design

The motivation behind the creation of the COLD database was the need for a comprehensive set of visual data that could be used to benchmark vision-based place classification algorithms for the purpose of localization and extraction of semantic information in an artificial, mobile cognitive system. An important property of such algorithms is robustness to variations that might occur in real-world environments. These include illumination variations as well as changes introduced by human activity in the environment (people appearing in the rooms, objects and furniture being relocated or removed). Robustness to categorical changes is another open issue in visual recognition. Humans are able to semantically label a room as "an office", "a kitchen" or "a corridor", even if they see it for the first time. This is because they are able to build robust categorical models of places. Providing similar capability for artificial systems is an extremely difficult task due to great within-category variability.

The aforementioned properties were reflected in the design of the COLD database to make it applicable in several different scenarios such as topological localization and mapping, semantic space labeling and metric mapping. Since the environments used for acquisition were located in different cities or even countries, they differed greatly with respect to spatial organization, appearance or imaging conditions. At the same time, as they served similar purpose, rooms of matching functionality could be found at all three sites. For the data acquisition, we tried to select rooms that are common to most of modern lab environments e.g a kitchen, a printer area or a corridor. However, some rooms were specific to particular labs, like the terminal room in Saarbrücken. The fact that, within the same environments and across the labs, there were several instances of rooms belonging to the same semantic category allows to use the database in the semantic space labeling scenario and provides sufficient within-class variability. At the same time, the stability of performance of localization algorithms can be tested in different settings. For each environment and room, the acquisition was repeated multiple times, over several days, under various illumination settings. As a result, the robustness of localization systems can be tested to dynamic appearance variations as well as occlusions introduced by human activity. Finally, we acquired dense sequences using two visual sensors and, when possible, a laser scanner. This makes the dataset useful for problems involving both vision and range sensors in both topological and metric mapping.

III. Acquisition Setup and Procedure

Three different mobile robots, the ActivMedia PeopleBot, the ActivMedia Pioneer-3 and the iRobot ATRV-Mini (see Fig. 1), were employed for image acquisition at the three labs. The PeopleBot and Pioneer-3 at Saarbrücken and

COLD-Saarbrücken COLD-Freiburg COLD-Ljubljana

Fig. 1. The three mobile robot platforms used for image acquisition.

Freiburg, were equipped with SICK laser scanners and wheel encoders whereas the iRobot at Ljubljana had only wheel encoders. In each case, the same camera setup was used for image acquisition. Two Videre Design MDCS2 digital cameras were used, one for perspective images and one for omnidirectional images. The catadioptric omnidirectional vision system was constructed using a hyperbolic mirror. The two cameras and the mirror were mounted together on a portable bracket as can be seen in Fig. 1. The heights of the cameras varied depending on the robot platform. All the images were acquired with the resolution of 640×480 pixels and the Bayer color pattern, with the auto-exposure mode turned on. The lens of the perspective camera provided the field of view of 84.9° x 68.9°.

The same procedure was followed during image acquisition at each lab. The robot was manually driven using a joystick (at a speed of roughly 0.3m/s) through each of the considered 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. At each lab, the acquisition was performed under several illumination settings (in cloudy weather, in sunny weather and at night) and at different times of day (during and after working hours) over a time span of two/three days. The acquisition was repeated at least thrice, resulting in a minimum of three image sequences, acquired one after the other, under similar conditions. Videos presenting the environment and the acquisition procedure in each lab are available as Ext. 4 (Freiburg), Ext. 5 (Ljubljana), and Ext. 6 (Saarbrücken).

At each lab, two different paths were followed by the robot during image acquisition: *(a)* the *standard* path, in case of which the robot was driven across rooms that are most likely to be found in most labs; *(b)* the *extended* path, in case of which the robot was additionally driven across the rooms that were specific for each lab. In Saarbrücken and Freiburg, the environments were divided into two parts (A and B), which were treated separately. As a result, two different sets of sequences were acquired. Tab. 1 provides a list of rooms in which the acquisition was performed and shows which rooms were included into which sequences for each lab. Ext. 1, 2 and 3 contain the maps of the laboratories in Freiburg,

Laboratory	Corridor	Terminal room	1-person office	2-persons office	Conference room	Printer area	Kitchen	Bath room	Large office	Stairs area	Lab
Saarbrücken	aAbB	A	AbB	aA	A	aAbB	B	aAbB			A
Freiburg	aAb		Ab	aAb		aA	A	aAb	A	aAb	
Ljubljana	aA			aA		aA	A	aA			aA

'a' - standard sequence, part A; 'A' - extended sequence, part A; 'b' - standard sequence, part B; 'B' - extended sequence, part B

Tab. 1 A list of the types of rooms that were used at the three labs. The letters indicate the sequences in which the rooms were included.

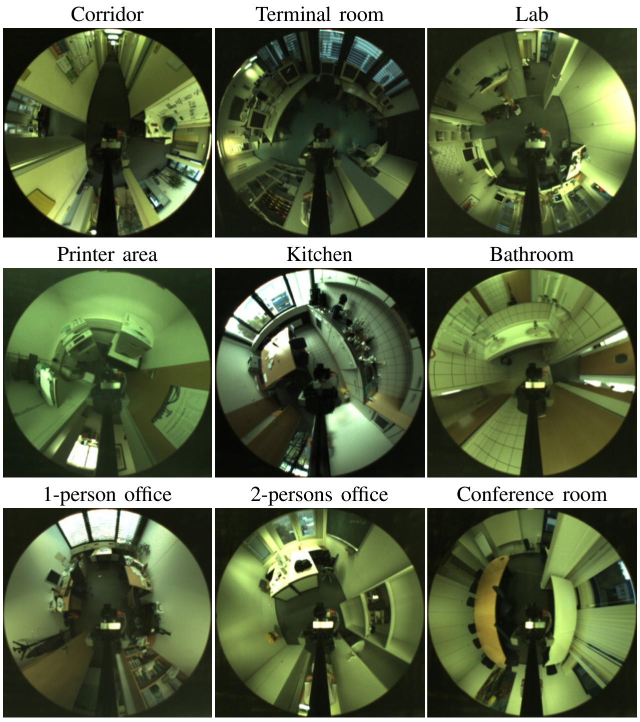


Fig. 3. Examples of omnidirectional camera images in COLD-Saarbrücken.

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	-	3
Freib.	3	3	3	-	4	3	3	-	3	-	4	-
Ljubl.	3	-	3	-	3	-	3	-	3	-	3	-

Tab. 2 Acquisition results for each of the three laboratories. The two different parts of the laboratories are annotated as 'A' and 'B'.

Ljubljana and Saarbrücken, respectively, with approximate paths followed by the robot during data acquisition.

IV. Acquisition Outcome

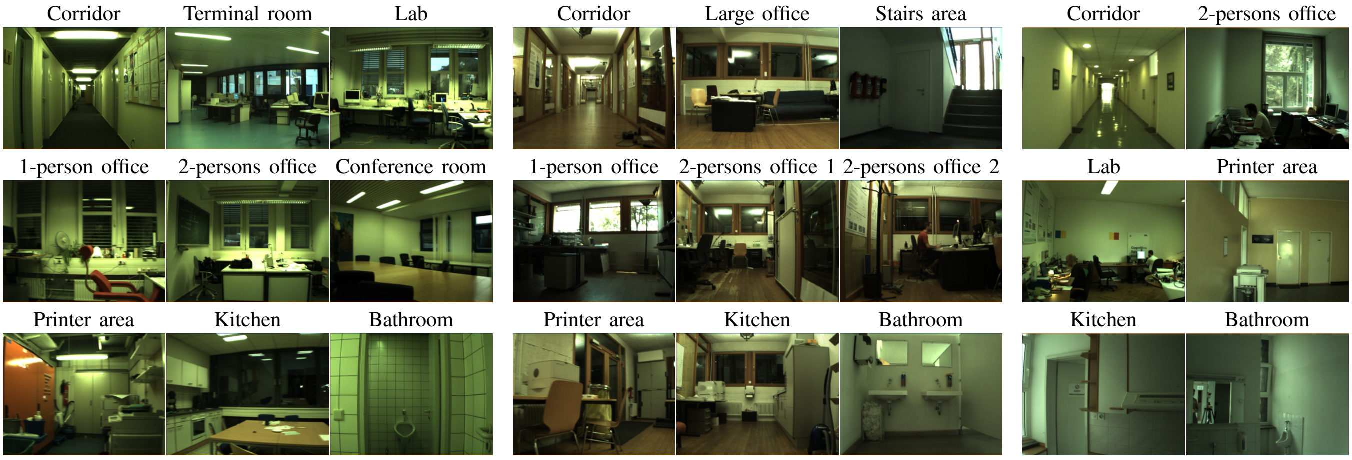
In total, 76 data sequences were acquired in 33 rooms belonging to 11 room categories. Detailed information about the number of sequences in the database for each lab, part and illumination setting can be found in Tab. 2. The total number of frames in each image sequence depends on the lab and the path that the robot followed (roughly 1000-2800 for Saarbrücken, 1600-2800 for Freiburg and 2000-2700 for Ljubljana). Fig. 2 presents examples of images acquired in each lab using the perspective camera and Fig. 3 shows omnidirectional images in COLD-Saarbrücken. Examples of omnidirectional images from the two other labs can be found in Ext. 7 and 8. Finally, videos presenting typical sequences in each sub-dataset can be found in Ext. 12-14 for perspective camera and in Ext. 15-17 for omnidirectional camera.

During image acquisition, special emphasis was placed on capturing natural variability that occurs in indoor environments which, in general, can be roughly categorized into dynamic, categorical and viewpoint variations. As can be seen in Fig. 4, the visual appearance of places varies in time because of human activity (furniture moved around, objects being taken in/out of drawers and etc., see Fig. 4a and Ext. 9) and illumination changes (day and night, artificial light on and off, see Fig. 4b and Ext. 10). These changes can be called dynamic because they are visible only when considering the environment across a span of time of at least several hours. Moreover, large within-category variability can be observed in the images (compare the room views in Fig. 2, see also Ext. 11). In case of Saarbrücken and Freiburg, the environments were divided into two parts, and rooms belonging to the same category can be found in both of them. As a result, we can distinguish between two levels of categorical variations: within one laboratory and across geographical locations. Finally, due to the manual control of the robot, differences in viewpoints occur between different sequences, even if they come from the same acquisition path. As the acquisition was performed continuously, and images were acquired also close to walls or furniture, some images contain little diagnostic information about the location.

V. Annotation and Processing

The images were annotated according to the following procedure: 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 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 COLD-Ljubljana, the annotation process was done using the odometry data with manual corrections. The poses for each image, 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.

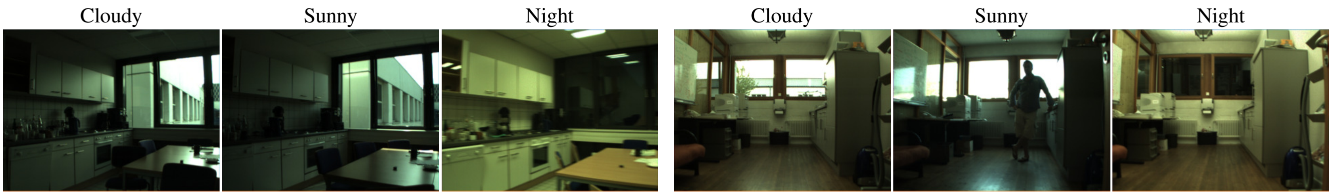
Additional tools and data files are available together with the database [8]. First, software is provided for unwrapping the omnidirectional images. Knowing the exact position of the mirror in the omnidirectional image is important for the unwrapping process, therefore data files with the



(a) Saarbrücken (b) Freiburg (c) Ljubljana

Fig. 2. Examples of perspective camera images in the database presenting the interiors of some of the rooms in each of the three labs.

(a) Variations introduced by human activity.



(b) Illumination variations.

Fig. 4. Images illustrating the dynamic visual variations introduced by human activity in the environment and changing illumination.

mirror center positions are provided as well. Finally, masks occluding the robot in the unwrapped images are available.

VI. Conclusions

We presented a database, called COsy Localization Database (COLD), consisting of data sequences acquired under varying conditions in three different laboratories across Europe using perspective and omnidirectional cameras mounted together on a socket. The database is applicable in several different scenarios such as robot localization or semantic labeling of space. The database is currently being expanded, and similar sequences have already been acquired at the Royal Institute of Technology in Stockholm, Sweden. These sequences will also be made publicly available.

The database was assessed in a set of baseline experiments with respect to both dynamic and categorical variations observed in the perspective images. The experiments were performed using a visual place classification algorithm based on local image features and support vector machines. For a detailed description of the method and obtained results, the reader is referred to [9].

References

[1] 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/

[2] "The PASCAL Visual Object Classes challenge." [Online]. Available: http://www.pascal-network.org/challenges/VOC/

[3] "The MIT-CSAIL database of objects and scenes." [Online]. Available: http://web.mit.edu/torralba/www/database.html

[4] A. Howard and N. Roy, "The Robotics Data Set Repository (Radish)," 2003. [Online]. Available: http://radish.sourceforge.net/

[5] E. Nebot, "The Sydney Victoria Park dataset." [Online]. Available: http://www-personal.acfr.usyd.edu.au/nebot/dataset.htm

[6] J. Luo, A. Pronobis, B. Caputo, and P. Jensfelt, "Incremental learning for place recognition in dynamic environments," in *Proceedings of the International Conference on Intelligent Robots and Systems (IROS07)*, San Diego, CA, USA, October 2007.

[7] "EU FP6 IST Cognitive Systems Integrated Project Cognitive Systems for Cognitive Assistants (CoSy)." [Online]. Available: http://www.cognitivesystems.org/

[8] M. M. Ullah, A. Pronobis, B. Caputo, J. Luo, and P. Jensfelt, "The COLD database," Kungliga Tekniska Hoegskolan, CVAP/CAS, Tech. Rep., October 2007. [Online]. Available: http://cogvis.nada.kth.se/COLD/

[9] M. M. Ullah, A. Pronobis, B. Caputo, J. Luo, P. Jensfelt, and H. I. Christensen, "Towards robust place recognition for robot localization," in *Proceedings of the International Conference on Robotics and Automation (ICRA08)*, Pasadena, CA, USA, May 2008.