The International Journal of Robotics Research

http://ijr.sagepub.com

COLD: The CoSy Localization Database

A. Pronobis and B. Caputo

The International Journal of Robotics Research 2009; 28; 588
DOI: 10.1177/0278364909103912

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

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

COLD: The CoSy Localization Database

Abstract

Two key competencies for mobile robotic systems are localization and semantic context interpretation. Recently, vision has become 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. In this paper we present 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.

KEY WORDS—vision-based robot localization, semantic place classification, robotic benchmarks and databases

1. Introduction

A major challenge to research on vision-based localization in mobile robotics is the difficulty of testing the robustness of algorithms in the presence of various visual variations. Since the results are highly dependent 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 make a fair comparison of solutions which are usually evaluated in different environments, under 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 (Griffin et al. 2007; Torralba et al. 2009; University of Southampton 2009). Also in robotics, research on simultaneous localization and mapping (SLAM) makes use of several publicly available datasets (Howard and Roy 2003; Nebot 2009). These are, however, primarily targeted for metric mapping problems and mostly contain odometry and range sensor data. A notable exception is the IDOL2 database (Luo et al. 2007), which can be seen as a preliminary attempt to create a database for visual place classification under dynamic changes.

This paper presents the CoSy (European Union 2009) 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 database, annotation files and tools for data processing are freely available at Pronobis (2008).

The COLD database consists of three independently collected sub-datasets, gathered over three distinct indoor laboratory environments containing spaces of common functionality,

The International Journal of Robotics Research
Vol. 28, No. 5, May 2009, pp. 588–594
DOI: 10.1177/0278364909103912

Figures 1–4 appear in color online: http://ijr.sagepub.com located in 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 refer to the three sub-datasets as COLD-Saarbrücken, COLD-Freiburg and COLD-Ljubljana.

The rest of the paper is organized as follows. In Section 2 we discuss further our motivations for building the database, the design and possible application scenarios. Sections 3 and 4 describe the acquisition setup, procedure and the acquisition outcome. The annotation and data post-processing are described in Section 5. We draw conclusions in Section 6. For further details on the database and a description of the baseline evaluation, we refer the reader to (Ullah et al. 2007, 2008).

2. 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 a 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 modern lab environments, e.g. a kitchen, a printer area, or a corridor. However, some rooms were specific to particular labs, such as 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 allowed us to use the database in the semantic space labeling scenario and provided sufficient within-class variability. At the same time, the stability of performance of the 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 to dynamic appearance variations can be tested, as can 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.

3. Acquisition Setup and Procedure

Three different mobile robots, the ActivMedia PeopleBot, the ActivMedia Pioneer-3 and the iRobot ATRV-Mini (see Figure 1(f)), were employed for image acquisition at the three labs. The PeopleBot and Pioneer-3 at Saarbrücken and 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 Figure 1(f). The heights of the cameras varied depending on the robot platform. All of 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 $84.9^\circ \times 68.9^\circ$.

The same procedure was followed during image acquisition at each lab. The robot was manually driven using a joystick (at a speed of roughly 0.2–0.4 m s^{-1}) through each of the considered rooms while continuously acquiring images at the rate of five frames per second. Since the two cameras were

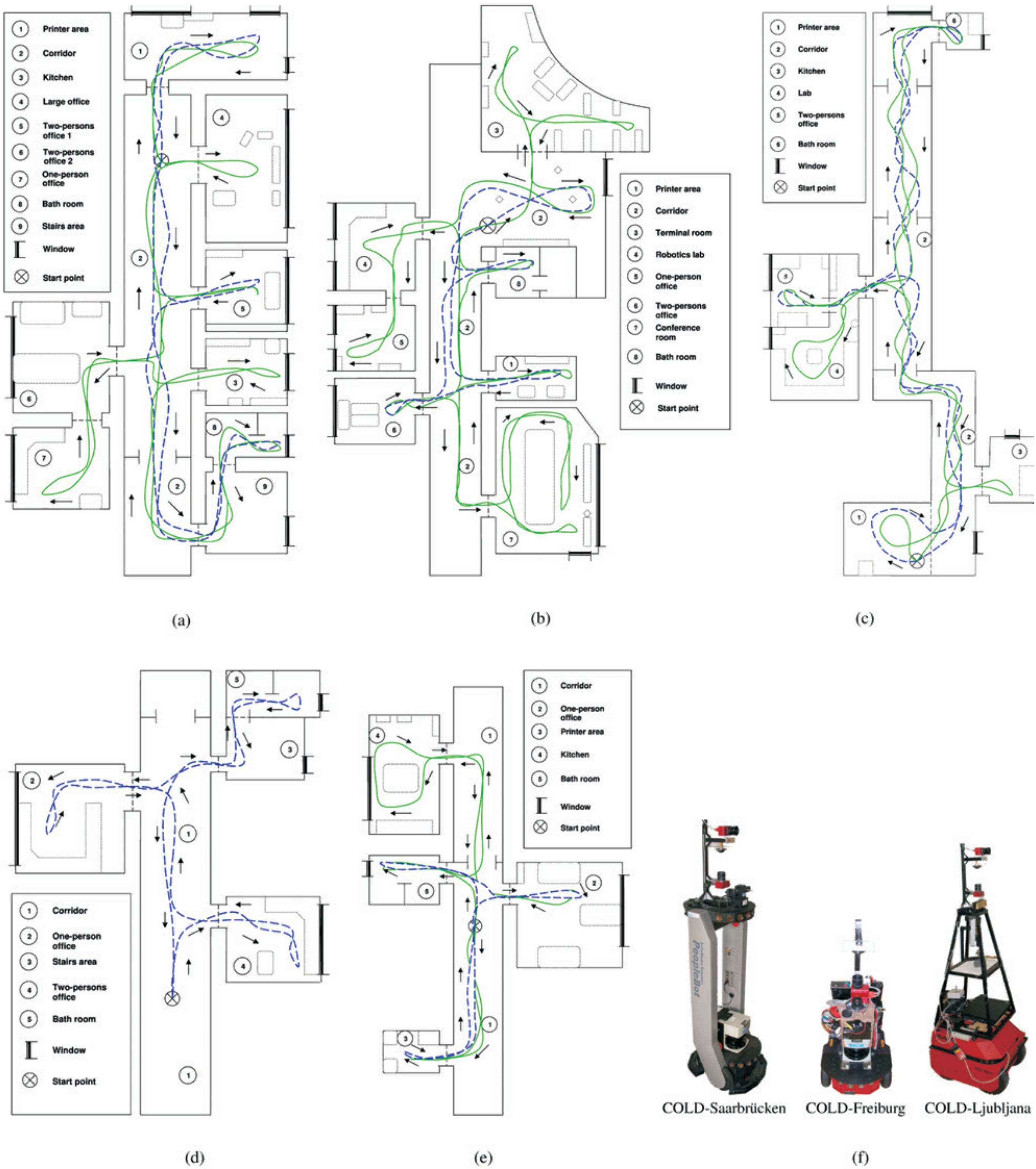


Fig. 1. Maps of the laboratories with approximate paths followed by the robots during data acquisition and the three mobile robot platforms: (a) Freiburg, Part A; (b) Saarbrücken, Part A; (c) Ljubljana; (d) Freiburg, Part B; (e) Saarbrücken, Part B; (f) robot platforms used for image acquisition. The standard path is plotted using a (blue) dashed line and the extended path is plotted using a (green) full line. Arrows indicate the direction in which the robot was driving.

Table 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.

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.

synchronized, for every perspective image, there is an omnidirectional image with the same timestamp. 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 three times, 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 on the website of the database.

At each lab, two different paths were followed by the robot during image acquisition: (a) the *standard* path, in which case the robot was driven across rooms that are most likely to be found in most labs; (b) the *extended* path, in which case 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. Figure 1(a)–(e) contain maps of the laboratories with approximate paths (both standard and extended) followed by the robot. Table 1 provides a list of rooms in which the acquisition was performed and shows which rooms were included into which sequences for each lab.

4. 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 Table 2(a). Average sequence parameters such as the total number of frames, the distance traveled by the robot, the time required to complete a sequence and the driving speed are given for each lab, part and sequence type in Table 2(b). Figure 2 presents examples of images acquired in each lab using the perspective camera and Figure 3 shows omnidirectional images in COLD-Saarbrücken. Examples of omnidirectional images from the two other labs can be found on the website of the database.

During image acquisition, special emphasis was placed on capturing the 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 Figure 4, the visual appearance of places varies in time because of human activity (furniture moved around, objects being taken in/out of drawers, etc.; see Figure 4(a)) and illumination changes (day and night, artificial light on and off, etc.; see Figure 4(b)). 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 Figure 2). In the 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, owing 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 also acquired close to walls or furniture, some images contain little diagnostic information about the location.

5. Annotation and Processing

The database was organized into a hierarchical directory structure according to the laboratory and part where the acquisition took place. The data gathered during each run were stored in a separate directory labeled with the sequence number and the type of illumination conditions during acquisition. For each sequence, 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 timestamp and 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 performed using the odometry data with manual corrections. For the perspective camera, an important consequence of this procedure is that the label assigned

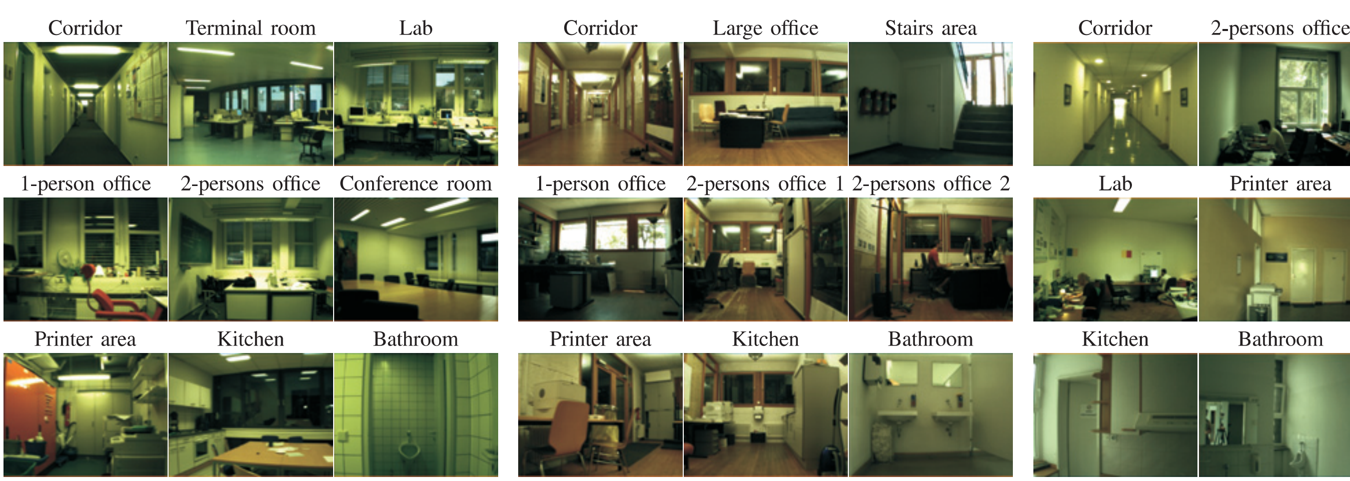
Table 2. Acquisition results for each of the three laboratories. The two different parts of the laboratories are annotated as "A" and "B". (a) Numbers of sequences for each location and sequence type. (b) Average sequence parameters for each location and sequence type.

(a)

	Standard sequences						Extended sequences					
	Cloudy		Night		Sunny		Cloudy		Night		Sunny	
Lab	A	B	A	B	A	B	A	B	A	B	A	B
Saarbrücken	3	5	3	3	—	3	3	3	3	3	—	3
Freiburg	3	3	3	—	4	3	3	—	3	—	4	—
Ljubljana	3	—	3	—	3	—	3	—	3	—	3	—

(b)

Lab	Part	Sequence	Frames	Distance (m)	Time (s)	Speed (m s^{-1})
Saarbrücken	A	Standard	1,167	53.4	250.3	0.215
		Extended	2,795	156.6	584.5	0.269
	B	Standard	781	38.1	174.6	0.220
		Extended	1,077	57.7	235.1	0.247
Freiburg	A	Standard	1,660	70.0	395.3	0.188
		Extended	2,547	104.2	572.8	0.185
	B	Standard	1,986	59.0	411.8	0.145
Ljubljana	A	Standard	2,153	180.0	452.9	0.393
		Extended	2,406	200.5	498.5	0.440



(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.

to a frame might be weakly related to its visual content owing to the constrained field of view.

The timestamp and pose for every image is encoded in its file name. Since the database can be labeled according to dif-

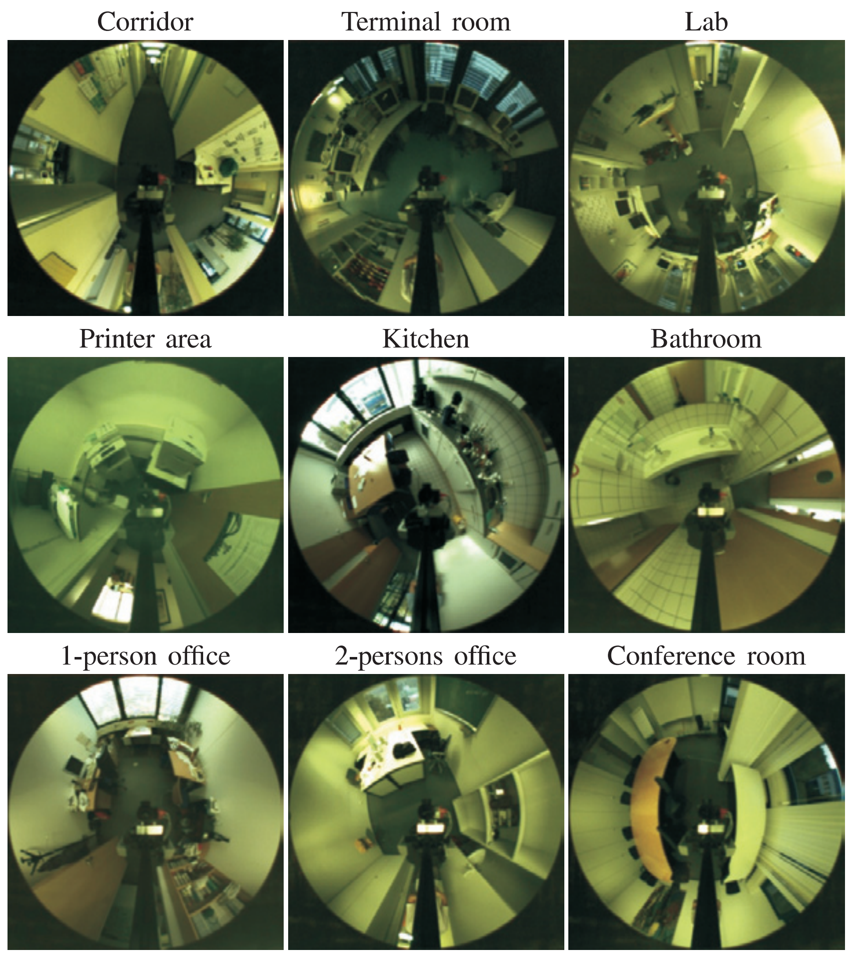
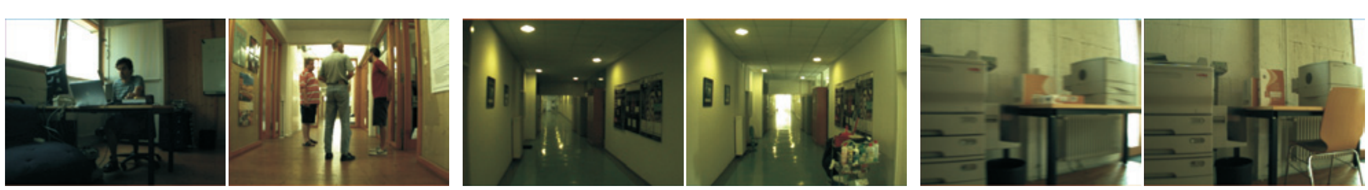
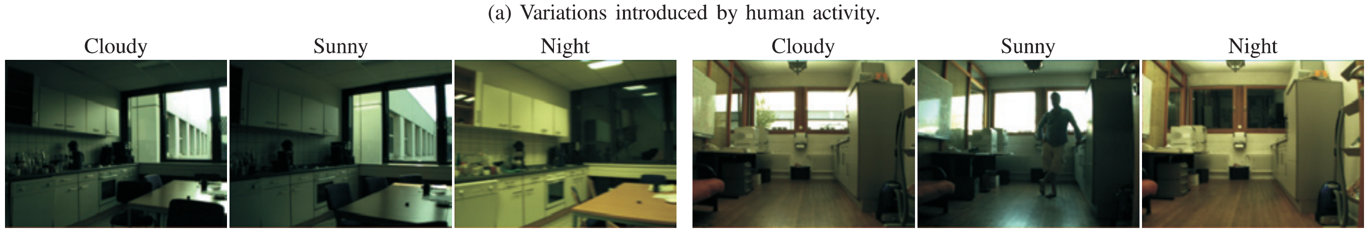


Fig. 3. Examples of omnidirectional camera images in COLD-Saarbrücken.



(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.

ferent strategies, room labels for each sequence are provided separately in a text file (each line contains an image file name and a label). Laser scans and odometry data are also stored in separate files. Additional tools and data files can be found on the website of the database (Pronobis 2008). First, scripts are available for accessing data gathered at a given location, under specified conditions and with a given label. Moreover, 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 mirror center positions are provided as well. Finally, masks occluding the robot in the unwrapped images are also available.

6. Conclusions

We have presented a database, called the 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 (Pronobis 2008). 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 has been 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 (Ullah et al. 2008).

Acknowledgements

This work was supported by the EU integrated projects CoSy FP6-004250-IP, CogX ICT-215181 (AP) and DIRAC IST-027787 (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 the acquisition of the COLD database possible: 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.

References

European Union (2009). EU FP6 IST Cognitive Systems Integrated Project Cognitive Systems for Cognitive Assistants (CoSy). http://www.cognitivesystems.org/.

Griffin, G., Holub, A., and Perona, P. (2007). Caltech-256 Object Category Dataset. *Technical Report 7694*, Caltech. http://authors.library.caltech.edu/7694/.

Howard, A. and Roy, N. (2003). The Robotics Data Set Repository (Radish). http://radish.sourceforge.net/.

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

Nebot, E. (2009) The Sydney Victoria Park dataset. http://www-personal.acfr.usyd.edu.au/nebot/dataset.htm.

Pronobis, A. (2008) The COLD Database. http://www.cas.kth.se/COLD/

Torralba, A., Murphy, K. P., and Freeman, W. T. (2009). The MIT-CSAIL database of objects and scenes. http://web.mit.edu/torralba/www/database.html.

Ullah, M. M., Pronobis, A., Caputo, B., Luo, J., and Jensfelt, P. (2007). The COLD database. *Technical Report*, Kungliga Tekniska Hoegskolan, CVAP/CAS, October.

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

University of Southampton (2009). The PASCAL Visual Object Classes challenge. http://www.pascal-network.org/challenges/VOC/.