KTH Computer Science
and Communication

TRITA-CSC-CV 2007:1 | ISRN KTH/CSC/CV--07/01--SE | ISSN 1653-6622

TECHNICAL REPORT

The COLD Database

**Muhammad Muneeb Ullah, Andrzej Pronobis,
Barbara Caputo, Jie Luo, Patric Jensfelt**

Stockholm 2007
CAS – Centre for Autonomous Systems
School of Computer Sciense and Communication
KTH Royal Institute of Technology

**Muhammad Muneeb Ullah, Andrzej Pronobis,
Barbara Caputo, Jie Luo, Patric Jensfelt**
CAS – Centre for Autonomous Systems

"The COLD Database"

Publicationsdate: October 2007
E-mail to author: pronobis@nada.kth.se
TRITA-CSC-CV 2007:1
ISRN KTH/CSC/CV--07/01--SE
ISSN 1653-6622

NADA is a co-operating departement between the
Royal Institute of Technology and Stockholm University

Reports can be ordered from:

Research Group:
CAS – Centre for Autonomous Systems
URL: www.csc.kth.se/cas

School of Computer Science and Communication
KTH (Royal Institute of Technology)
SE- 100 44 Stockhom, Sweden

URL: www.csc.kth.se

The COLD Database [1]

Muhammad Muneeb Ullah[⋆‡], Andrzej Pronobis[‡], Barbara Caputo[⋆†],
Jie Luo[⋆†], and Patric Jensfelt[‡]

[⋆]IDIAP Research Institute
1920 Martigny, Switzerland
[†]EPFL, 1015 Lausanne, Switzerland
[mullah,bcaputo,jluo]@idiap.ch

[‡]Centre for Autonomous Systems
Royal Institute of Technology
SE-100 44 Stockholm, Sweden
[pronobis, patric]@kth.se

October 16, 2007

[1]This work was supported by the EU integrated project CoSy FP6-004250-IP, www.cognitivesystems.org, and the Swedish Research Council contract 2005-3600-Complex. The support is gratefully acknowledged. Very special thanks go to Henrik I. Christensen and the 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.

1 Introduction

This report provides a detailed description of the COLD database. The acronym COLD stands for COsy[1] Localization Database. The database represents an effort to provide a large-scale, flexible testing environment for evaluating mainly vision-based localization systems aiming to work on mobile platforms in realistic settings. The COLD database consists of three separate 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 Systems Laboratory at the University of Freiburg, Germany; and the Language Technology Laboratory at the German Research Center for Artificial Intelligence in Saarbrücken, Germany.

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 localization as well as place recognition and categorization algorithms. An important property of such algorithms is a certain level of robustness against different kinds of variations that might occur in real-world environments. From the point of view of vision-based solutions, it is crucial to provide invariance with respect to dynamic changes influencing the appearance of places observed over time. These are mainly 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). Invariance to categorical changes is another open issue in visual recognition. Humans are able to categorize 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 an artificial place categorization system is an extremely difficult task due to great within-category variability. Finally, a universal localization system should deliver constant performance independently of the environment it is applied in.

A major obstacle for the development of localization systems robust with respect to a broad range of variability is the lack of a unified testbed for benchmarking and comparing competing solutions. The COLD database tries to fill the gap by providing a large versatile set of image sequences acquired at three different laboratory environments under natural variations. The sequences in the database were recorded using several mobile robot platforms 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 perspective and omnidirectional cameras were mounted on a portable mounting 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 weather and illumination conditions and across a time span of two/three days. Special care was taken in the choice of the rooms to image. For each laboratory, there exists a set of sequences containing rooms of similar functionalities that can also be found in the other two. Thus, the COLD database is an ideal testbed for assessing the robustness of localization and recognition algorithms with respect to both dynamic and categorical changes. In this report, we will refer to the three databases with the names of the cities where the acquisition took place (COLD-Saarbrücken, COLD-Freiburg and COLD-Ljubljana).

The rest of the report is organized as follows: Section 2 describes the indoor environments at the three laboratories and Section 3 provides information about the robot platforms and the cameras employed during acquisition. Then, Section 4 explains the acquisition procedure that was followed and Section 5 describes the technique used to annotate the images sequences. The different types of variability captured in the image sequences are illustrated in Section 6 and Section 7 sheds some light on the difficulties that were incurred during and after the image acquisition. Finally, Section 8 presents several possible future extensions to the database. Detailed information about the coordinate system, data format, and structure of the database can be found in the appendices.

2 The Environments

As already mentioned, the database was acquired in three different indoor office/laboratory environments in Freiburg and Saarbrücken in Germany as well as in Ljubljana, Slovenia. Since the environments were located in different cities or even countries, they differed greatly with respect to spatial organization, appearance or imaging conditions. At the same time, due to the fact that they served similar purpose,

[1] EU FP6 IST Cognitive Systems Integrated Project Cognitive Systems for Cognitive Assistants (CoSy) (FP6-004250-IP).

LAB	Corridor	Terminal room	Robotics lab	1-person office	2-persons office	Conference room	Printer area	Kitchen	Bath room	Large office	Stairs area	Lab
Saarb.	√	√	√	√	√	√	√	√	√	-	-	-
Freib.	√	-	-	√	√	-	√	√	√	√	√	-
Ljubl.	√	-	-	-	√	-	√	√	√	-	-	√

Table 1: List of types of rooms that were used during image acquisition at each of the three labs.

Robot platform	ActivMedia PeopleBot Saarbrücken		ActivMedia Pioneer-3 Freiburg		iRobot ATRV-Mini Ljubljana	
Camera type	Perspective	Omni	Perspective	Omni	Perspective	Omni
Frame rate	5 fps					
Resolution	640×480 pixels, Bayer color pattern					
Exposure	Automatic					
Field of view	68.9° × 54.4°	—	68.9° × 54.4°	—	68.9° × 54.4°	—
Height	140cm	116cm	66cm	91cm	159cm	153cm

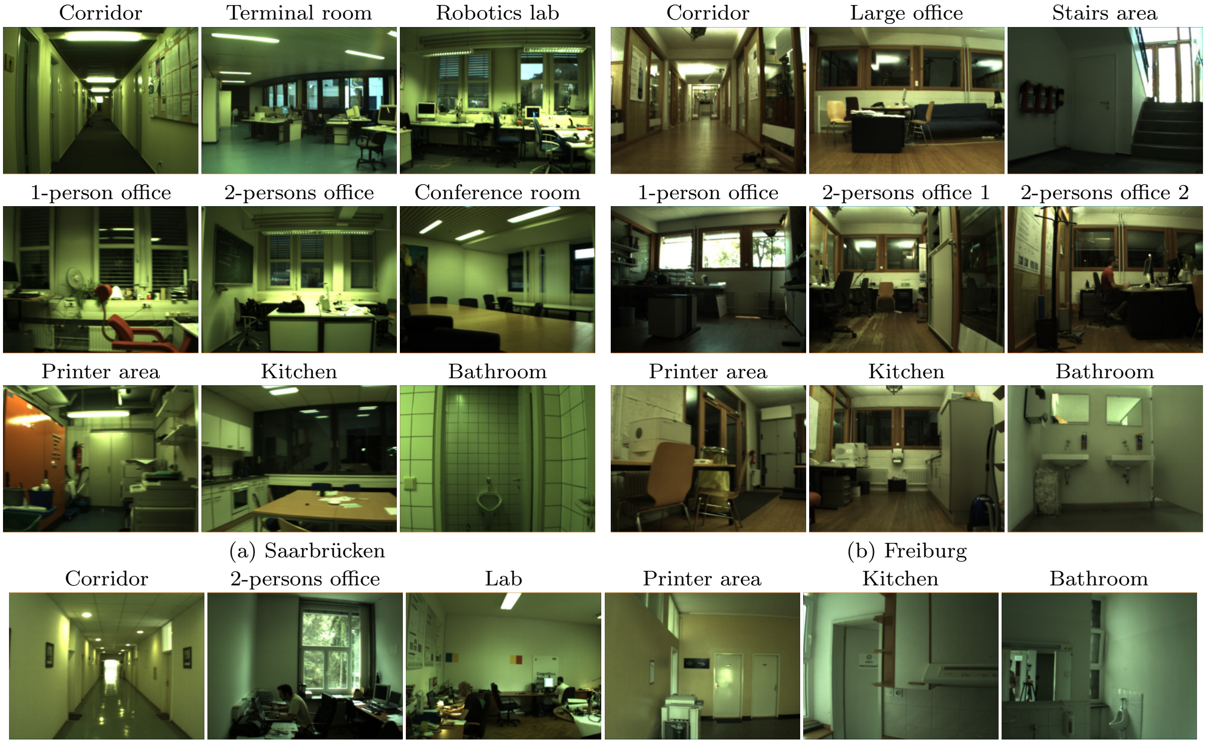
Table 2: Parameters and settings of the cameras for each robot platform.

rooms of matching functionality such as offices, corridors or bathrooms could be found at all three sites. For image acquisition, usually rooms that are common for most modern laboratory environments were selected; however, in some cases, rooms specific to particular site were also used e.g the terminal room in Saarbrücken. Table 1 provides a list of types of rooms that were used during acquisition. Sample images showing the interior of each room captured with both perspective and omnidirectional cameras are shown in Figure 1-2. General maps of the indoor environments are presented in Figure 4-5. In case of Saarbrücken and Freiburg, two different parts of the laboratories were considered separately (referred to as part A and B) and can be seen as two similar individual environments.

In each of the three cases, only subsections of the whole environment were selected for image acquisition. The selected rooms fulfill different purposes and that determines the human activity that is likely to occur and the way furniture is set up or objects are placed. For instance, furniture organization in an office, printer area or a kitchen will be totally different. Rooms like a corridor or a printer area can be regarded as stable as the furniture is mostly fixed and objects are less moveable. Consequently, these rooms were less susceptible to variations caused by human activity compared to rooms like an office or a kitchen, where the furniture (e.g. chairs) is moved more often, objects (e.g. cups) are displaced frequently and the decoration is changed by the owners. However, since the span of time for image acquisition at each of the environments was only two/three days, no drastic variations due to decoration change were observed. Additionally, rooms like a kitchen, corridor or a printer area can be seen as public, which means that various people may be present. At the same time, the offices and the laboratory rooms were usually imaged empty or with their owners at work. Finally, the visual appearance of different rooms (especially the regions close to windows) was heavily affected by illumination and weather conditions. For instance, the image sequences acquired under sunny weather capture shadows and reflections caused by direct sunlight. The imaging conditions differed for each of the three laboratories and the image set recorded in Freiburg was particularly affected by variations in illumination due to large number of windows and the fact that rooms were physically separated by glass walls and doors. On the other hand, in case of Ljubljana, some places were weakly illuminated (especially at night), which introduced additional noise during acquisition.

3 Robot Platforms and Camera Setup

Three different mobile robot platforms, the ActivMedia PeopleBot, the ActivMedia Pioneer-3 and the iRobot ATRV-Mini (see Figure 3), 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 was equipped only with the wheel encoders. At each lab, the robot was manually controlled using a joystick.



(c) Ljubljana

Figure 1: Exemplary images acquired using the perspective camera showing the interiors of the rooms at the three laboratories.

LAB	Standard sequences						Extended sequences					
	Cloudy		Night		Sunny		Cloudy		Night		Sunny	
	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	-

Table 3: Numbers of sequences acquired at each of the three laboratories. Two different parts of the laboratories are denoted as ‘A’ and ‘B’.

The camera setup was built using two Videre Design[1] MDCS2 digital cameras; one for perspective images, and one for the 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 presented in Figure 3. The bracket was then moved from one lab to the other, and mounted on the robot platform available at each site. The heights at which the cameras were mounted differed and depended on the height of the robot platform used for acquisition. The same settings of the cameras were used at the three labs. Detailed parameters and settings of the cameras for each platform can be found in Table 2.

4 Acquisition Procedure

The same procedure was followed during image acquisition at each lab. The robot was manually driven (at a speed of roughly 0.3 m/s) through each of the 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. The acquisition was performed under various

[1] www.videredesign.com

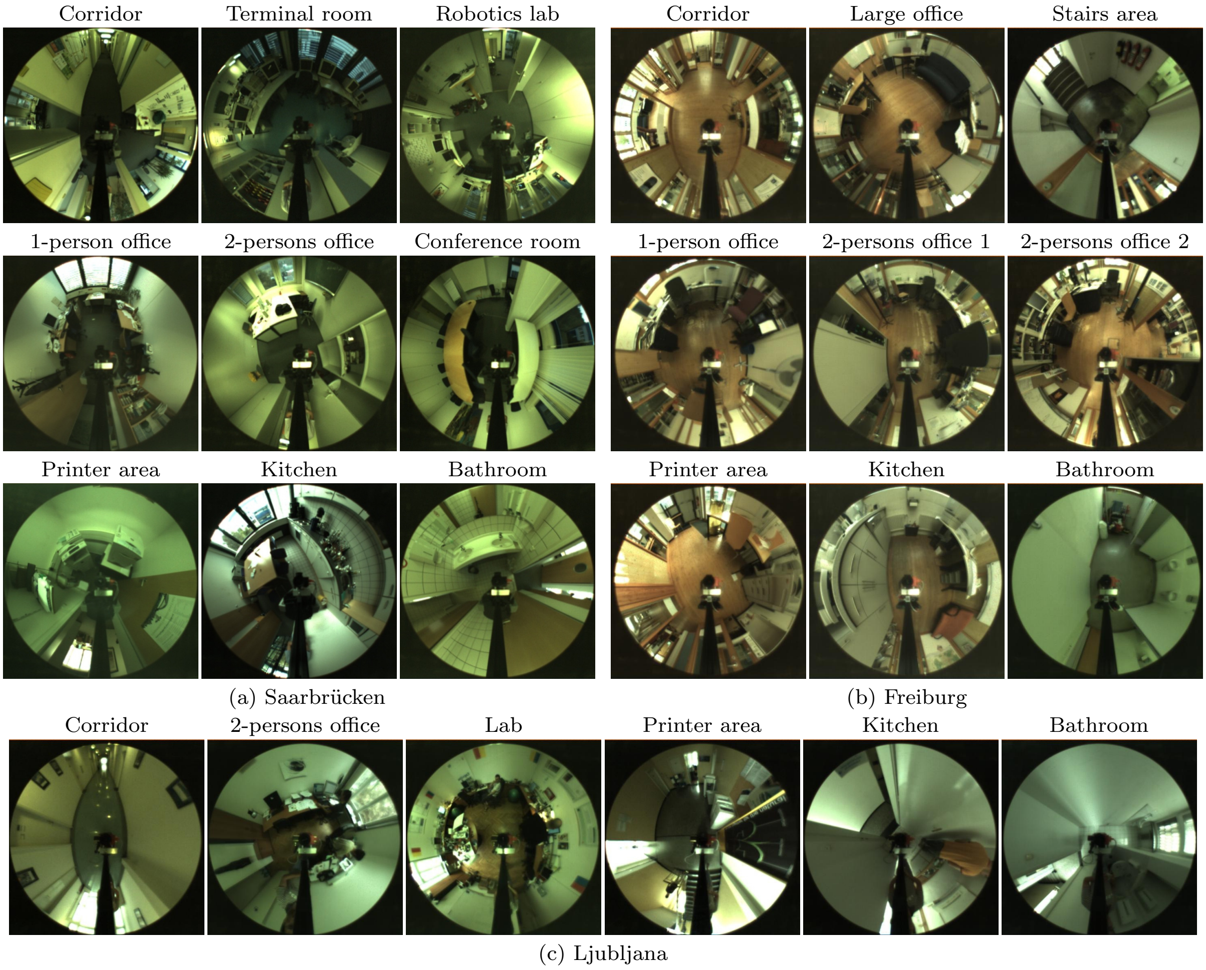


Figure 2: Exemplary images acquired using the omnidirectional camera in each of the rooms at the three laboratories.

LAB	Corridor	Terminal room	Robotics lab	1-person office	2-persons office	Conference room	Printer area	Kitchen	Bath room	Large office	Stairs area	Lab
Saarb.	XOΔ□	Δ	Δ	OΔ□	XΔ	Δ	XOΔ□	□	XOΔ□	-	-	-
Freib.	XOΔ	-	-	OΔ	XOΔ	-	XΔ	Δ	XOΔ	Δ	XOΔ	-
Ljubl.	XΔ	-	-	-	XΔ	-	XΔ	Δ	XΔ	-	-	XΔ

Table 4: List of rooms covered by each type of sequence at each lab. Each room is marked with different shapes according to the sequences in which it has been included: 'X' stands for standard sequence A; 'O' stands for standard sequence B; 'Δ' stands for extended sequence A; and '□' stands for extended sequence B.

ActivMedia PeopleBot at Saarbrücken — ActivMedia Pioneer-3 at Freiburg — iRobot ATRV-Mini at Ljubljana

Figure 3: Three different mobile platforms employed for image acquisition at the three labs. The portable socket with the camera setup is shown in the top right corner.

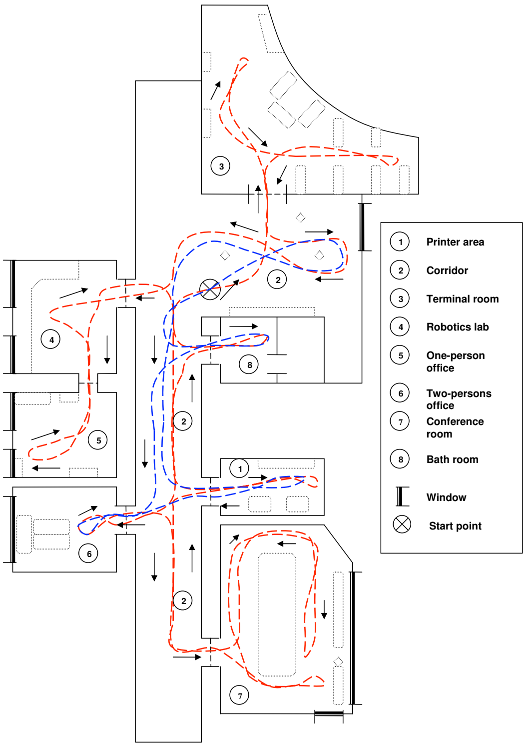
illumination and weather conditions that could be classified into three groups: sunny weather, cloudy weather and night. For the different illumination conditions, the acquisition procedure was repeated at least thrice, resulting in a minimum of three image sequences, acquired one after the other, under similar conditions.

At each lab, two different paths were followed by the robot during image acquisition: *(a)* the *standard* path, where the robot was driven across rooms that are most likely to be found in most labs; *(b)* the *extended* path, where the robot was driven across all the available rooms. Figure 4-5 presents the two types of paths that the robot followed in each environment. The extended path generally contained more rooms than the standard path, and the additional rooms are usually specific for each particular lab. In each case, a set of standard and extended image sequences was collected at each lab. Detailed information about the number of sequences in the database can be found in Table 3. Table 4 presents a list of rooms covered by each sequence type at each lab. Due to the manual control of the robot, strong viewpoint variations can be observed between different sequences, even if they were recorded following the same type of acquisition path. 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).

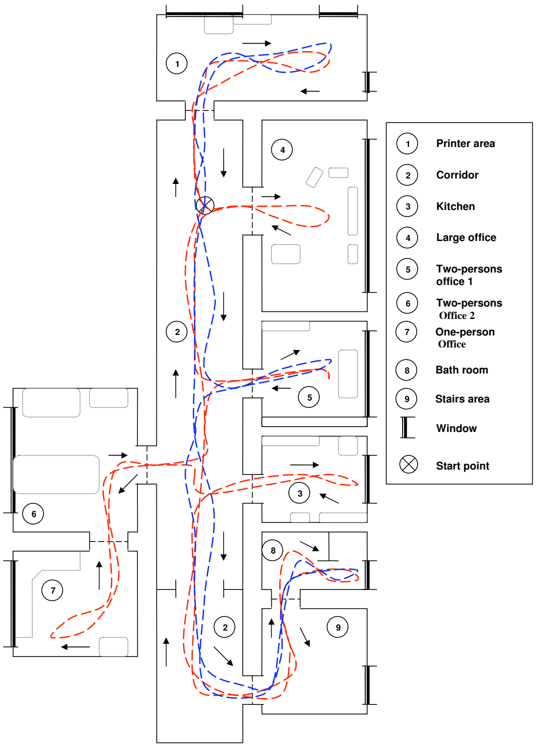
5 Data Annotation

In order to label the acquired images, the same procedure as in [3, 2] was followed: the robot pose was estimated during the acquisition process using a laser-based localization technique [1]. The pose was represented in a predefined global coordinate system (see Appendix C). Each image was then labeled as belonging to one of the available rooms according to the position (i.e. estimated coordinates in the global coordinate system) of the robot at the moment of acquisition. This strategy could not be directly followed in Ljubljana, because the robot patform did not have a laser scanner. Thus, for the sequences captured in Ljubljana, the annotation procedure was accomplished using odometry data with manual corrections. Description of the file format used to store odometry and laser range data can be found in Appendix B.

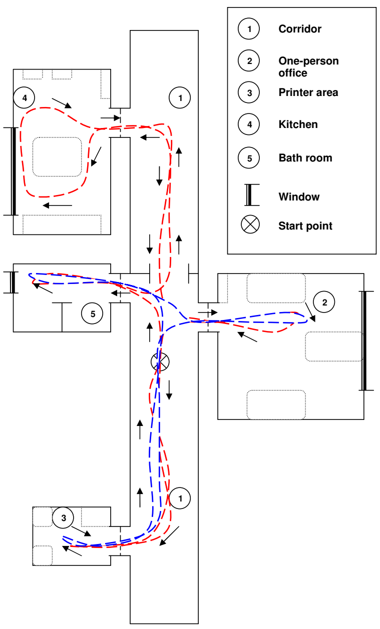
For the perspective camera, an important consequence of this annotation procedure is that when the



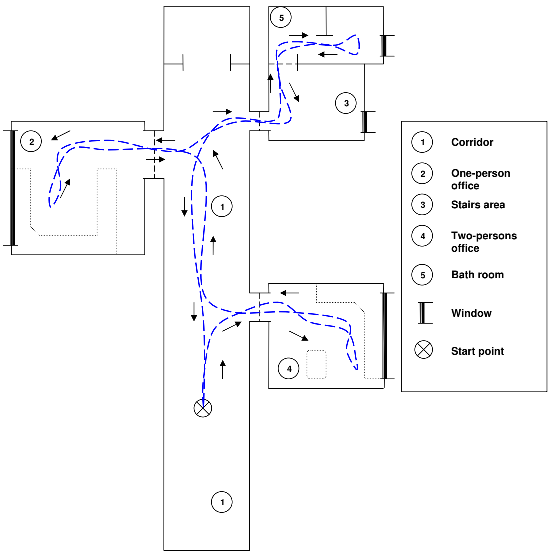
(a) Map of Saarbrücken Portion A.



(b) Map of Freiburg Portion A.



(c) Map of Saarbrücken Portion B.



(d) Map of Freiburg Portion B.

Figure 4: General maps of the laboratories in Saarbrücken and Freiburg with two different paths followed by the robot. The standard path is presented using blue dashed line and the extended path is marked using red dashed line. Arrows indicate the direction of driving of the robot. Boundaries between rooms are marked with gray dashed lines and approximate outline of furniture is plotted with dotted lines.

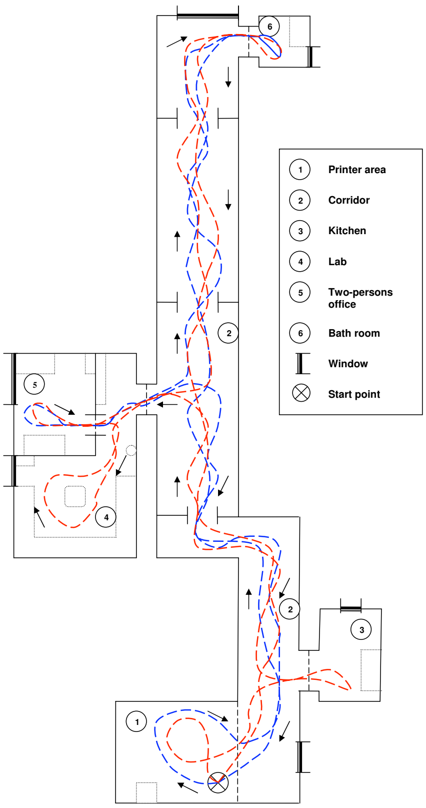


Figure 5: General map of the laboratory in Ljubljana with two different paths followed by the robot. The standard path is presented using blue dashed line and the extended path is marked using red dashed line. Arrows indicate the direction of driving of the robot. Boundaries between rooms are marked with gray dashed lines and approximate outline of furniture is plotted with dotted lines.

(a) Corridor (b) Kitchen

Figure 6: Influence of illumination variations on the visual appearance of the environments.

robot was positioned in the transition region between two rooms, the label assigned to a frame might be weakly related to its visual content. 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 dataset will be particularly challenging.

6 Captured Variability

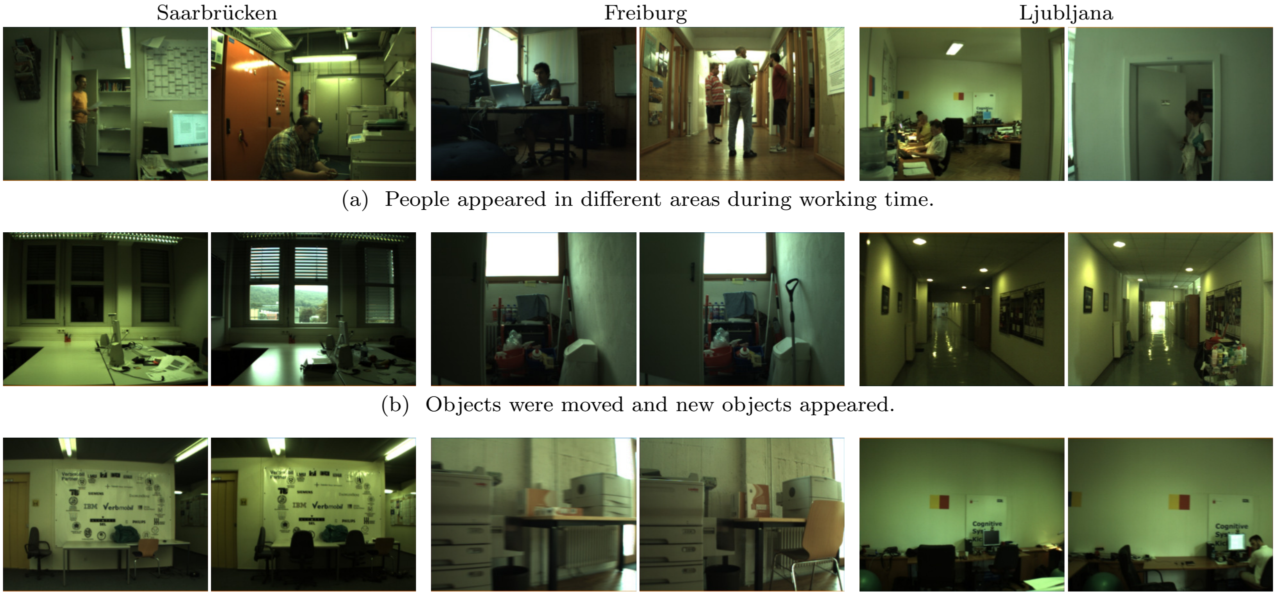
During image acquisition, special emphasis was placed on capturing natural variability that occurs in indoor environments. For this reason, all the image sequences at each laboratory were acquired during the working time, over a time span of two/three days, and under different illumination and weather conditions. As a result, different visual variations can be observed in the recorded image sequences. In general, the types of variability captured in the three environments can be roughly categorized into dynamic, categorical and viewpoint variations. The following subsections discuss each type in detail.

6.1 Dynamic Variations

The visual appearance of places varies in time because of illumination changes (day and night, artificial light on and off) and human activity (furniture moved around, objects being taken in/out of drawers and etc.). These changes can be called dynamic because they are visible just when considering the indoor environment across a span of time of at least several hours.

Illumination variations had a very significant influence on the visual appearance of the indoor environment at each lab. In order to capture this variability, image sequences were acquired under three different illumination and weather conditions: *cloudy* weather, *sunny* weather and *night*. The influence of illumination can be observed in the exemplary images presented in Figure 6. Each type of weather and illumination conditions had a different effect on the visual appearance of the indoor environments due to different amounts of natural and artificial light available in each case. For sunny weather, the natural sunlight dominated and introduced shadows and reflections, while at night only artificial light was available. Moreover, the fact that the auto-exposure mode was on resulted in lower contrast in the informative parts of images when the camera was directed towards a bright window. The illumination conditions during the cloudy weather can be regarded as intermediate between those during sunny day and those at night. Apart from illumination changes, the overall visual appearance of the indoor environments was changing over time as they were being used by the humans. The following variations introduced by human activity can be observed in the images:

- People appeared in different parts of the indoor environment during the office hours (Figure 7a).



(a) People appeared in different areas during working time.

(b) Objects were moved and new objects appeared.

(c) Pieces of furniture were moved around.

Figure 7: Images illustrating the influence of human activity on the indoor environments.

- Objects were moved and taken in/out of the cupboards/drawers (Figure 7b).
- Pieces of furniture, such as chairs, were moved around (Figure 7c).

6.2 Categorical Variations

The COLD database was acquired in several environments consisting of rooms that can be assigned to the same functional category. As a result, large within-category variability can be observed in the images. In case of the laboratories in 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. Figures 9 and 8 illustrate the categorical variations captured in the database.

6.3 View-point Variations

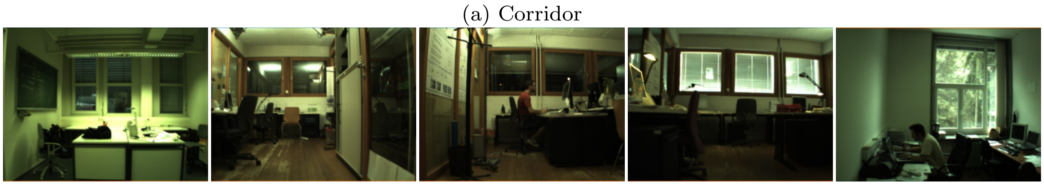
Similar, roughly planned paths were followed while driving the robots during image acquisition. Still, mainly due to the manual control of the robots, view-point variations occurred between the image sequences.

7 Difficult Examples

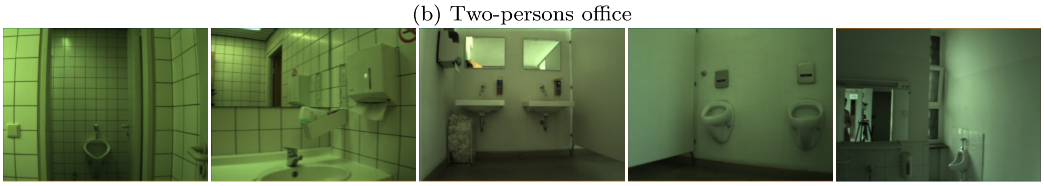
As already mentioned, the image annotation was based on the position of the robot during acquisition rather than the contents of the images. As a result, in case of perspective images, the labels might not be consistent with the visual information when the robot was positioned in a transition region between two rooms. The situation is illustrated in Figure 10. The information captured in omnidirectional images can be used to mitigate this problem. Another difficulty that a place recognition system may encounter comes from the fact that images were acquired also close to walls or furniture. As a result, the database contains images with little diagnostic information about the places were the acquisition was performed. Figure 11 presents a few non-informative images that can be found in the COLD database. Finally, the motion of the robots introduced additional distortions and some of the images might be affected by motion blur. All these factors make the database suitable for testing system that are meant to be applied on mobile platforms where such problems are common.

Figure 8: Images of rooms belonging to the same functional category captured in the three laboratories.

(a) Corridor



(b) Two-persons office



(c) Bathroom

Figure 9: The within-category variability of the visual appearance of three types of rooms captured in the COLD database.

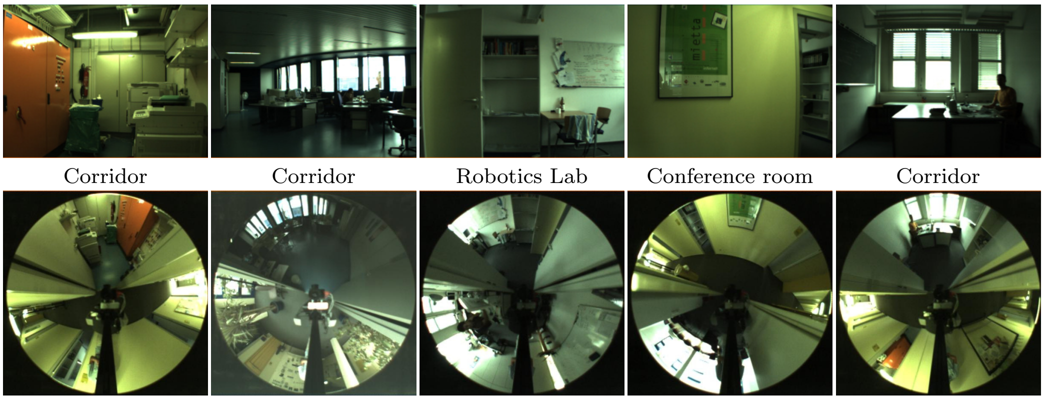


Figure 10: Exemplary images from the COLD-Saarbrücken dataset illustrating the limitations of the labelling technique. The figure shows images acquired with perspective camera with labels assigned on the basis of the location of the robot. The labels do not correspond with the visual information in the images due to the relatively narrow field of view of the cameras.

Figure 11: Examples of non-informative images in the COLD database.

8 Future Extensions to the Database

Several extensions to the database are planned in the future. Currently, fourth set of image sequences at the Computer Vision and Active Perception Laboratory in Stockholm is being acquired. Moreover, the span of time during image acquisition for the COLD database was just two/three days. Therefore, it was not possible to capture long-time variations that occur in indoor environments. As a result, it would be of interest to repeat the acquisition after some time.

References

[1] J. Folkesson, P. Jensfelt, and H. Christensen. "Vision SLAM in the Measurement Space". In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'05)*, Barcelona, Spain, April 2005.

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

[3] A. Pronobis, B. Caputo, P. Jensfelt, and H. I. Christensen. A discriminative approach to robust visual place recognition. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS06)*, Beijing, China, October 2006.

A File Naming Conventions

Complete information about each image is encoded in its filename. The naming convention used to generate the image filenames, written in the Backus Naur Form notation, is presented below:

t⟨*timestamp*⟩_x⟨*x.coordinate*⟩_y⟨*y.coordinate*⟩_a⟨*angle*⟩.⟨*extension*⟩

⟨*timestamp*⟩ - The internal epoch time of the robot at which the image was taken.

⟨*x.coordinate*⟩ - X-coordinate specifying position of the robot in the global coordinate system (unit:m).

⟨*y.coordinate*⟩ - Y-coordinate specifying position of the robot in the global coordinate system (unit:m).

⟨*angle*⟩ - Angle representing the pose of the robot (unit:rad).

⟨*extension*⟩ - File extension. Depends on the image format.

The internal clock of the robot was used in order to timestamp the images. For some sequences, there might be a few hours deviation with respect to the real time.

The information about the label of each of the acquired images is not included in the filename. However, a list of all the image file names together with the labels of the rooms where the images were acquired, is provided separately in a text file. Detailed description of the format of this file can be found in Appendix D.

B Odometry and Scans File Format

The odometry and laser range data collected during the image acquisition are also provided with the image sequences and are stored in the *odom.tdf* and *scans.tdf* files. The files use the format that is supported by the "CURE" toolbox package[2], which is a C++ based software library that provides utility algorithms for robot control. The formats of the odom.tdf and scans.tdf files are described below:

- **odom**:
  Ø Ø Ø time_sec time_usec Ø Ø Ø x y z theta phi psi

- **scans**:
  Ø Ø numPts time_sec time_usec Ø Ø Ø number_of_warningsflags number_of_intesitylevels scanner_type sensor_id angle_step start_angle max_range range_res range_1 range_2 .... range_N

For odometry file, only x, y, theta and the time are used and z, phi, psi will always be zero.

[2]http://www.cas.kth.se/CURE/

C Coordinate System Used In Labeling

Global coordinate system was used in order to represent the pose of the robots. Figures 12-13 present the maps of the three laboratories with the coordinate system. For labeling, the coordinates have been shifted from the center of the robots to the position of the cameras.

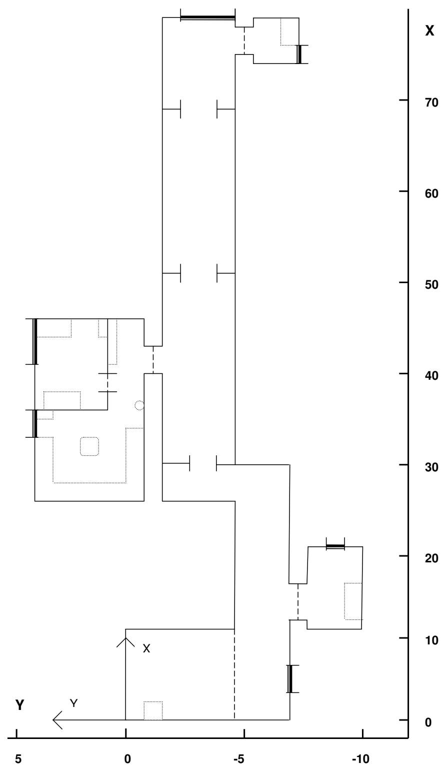


Figure 12: Coordinate system used for position estimation at Ljubljana.

(a) Map of Saarbruecken Portion A.

(b) Map of Freiburg Portion A.

(c) Map of Saarbruecken Portion B.

(d) Map of Freiburg Portion B.

Figure 13: Coordinate systems used for position estimation at Saarbruecken and Freiburg.

D Structure of COLD Database

This section provides information about the directory structure of the database. For the sake of simplicity, the COLD database was divided into subdirectories according to the laboratory where the acquisition was performed. The full directory structure is presented below:

```
⟨COLD⟩
|
|
+–COLD_<lab>                                  – Dataset
     |
     |
     +–Portion_<code>                         – Environment part
           |
           |
           +–<sequence>                       – Image sequence
                |
                |
                +–std_cam\                    – Directory containing perspective images
                |
                |
                +–omni_cam\                   – Directory containing omnidirectional images
                |
                |
                +–odom_scans\                 – Directory containing odom.tdf and scans.tdf files
                |
                |
                +–places\                     – Directory containing the labels.lst file
```

⟨*lab*⟩::=Saarbruecken|Freiburg|Ljubljana

⟨*code*⟩::=A|B

⟨*sequence*⟩::=seq<id>_<illumination><no>
⟨*id*⟩::=1|2|3|4
⟨*illumination*⟩::=cloudy|night|sunny
⟨*no*⟩::=1|2|3|4|5

The format of the *labels.lst* file is described below:

- **labels.lst:**

  ⟨*image_name*⟩ ⟨*room*⟩

  ⟨*image_name*⟩ - Line format is the same as given in Appendix A
  ⟨*room*⟩ - Label of the room where the picture was taken.

::=	CR-<code>	- Corridor	|	SA-<code>	- Stairs area
|	TR-<code>	- Terminal room	|	LAB-<code>	- Lab
|	RL-<code>	- Robotics lab			
|	1PO-<code>	- One-person office			
|	2PO|2PO1|2PO2-<code>	- Two-persons office			
|	CNR-<code>	- Conference room			
|	PA-<code>	- Printer area			
|	KT-<code>	- Kitchen			
|	BR-<code>	- Bathroom			
|	LO-<code>	- Large office			

CAS – Centre for Autonomous Systems
TRITA-CSC-CV 2007:1
ISRN KTH/CSC/CV--07/01--SE
ISSN 1653-6622

CAS – Centre for Autonomous Systems
School of Computer Science and Communication
KTH Royal Institute of Technology
www.csc.kth.se