The KTH-IDOL2 Database

Jie Luo, Andrzej Pronobis, Barbara Caputo, and Patric Jensfelt
Computational Vision and Active Perception Laboratory (CVAP)
School of Computer Science and Communication (CSC)
KTH, SE-100 44 Stockholm, Sweden
CVAP-Report 304
{jiel, pronobis, caputo, patric}@nada.kth.se

October 13, 2006

1 Introduction

This report provides a detailed description of the KTH-IDOL2 database. The name IDOL is an acronym which stands for Image Datebase for rObot Localization. The database was created for the purpose of evaluating the robustness and adaptability of a vision-based place recognition algorithms to changes that occur in real-world dynamic environments.

The database contains 24 image sequences accompanied by laser scans and odometry data acqired using two mobile robot platforms. The acquisition was performed within an indoor laboratory environment consisting of five rooms with different functionality (one-person office, two-persons office, corridor, kitchen, and printer area) under various illumination conditions (cloudy weather, sunny weather, and night) across a span of six months. As a result, the data capture natural variability that occur in real-world environments introduced by both illumination and human activity (people appear in the rooms, furniture and objects change location etc.).

The KTH-IDOL2 database is an extension of the KTH-IDOL1 database and as such consists of the following parts:

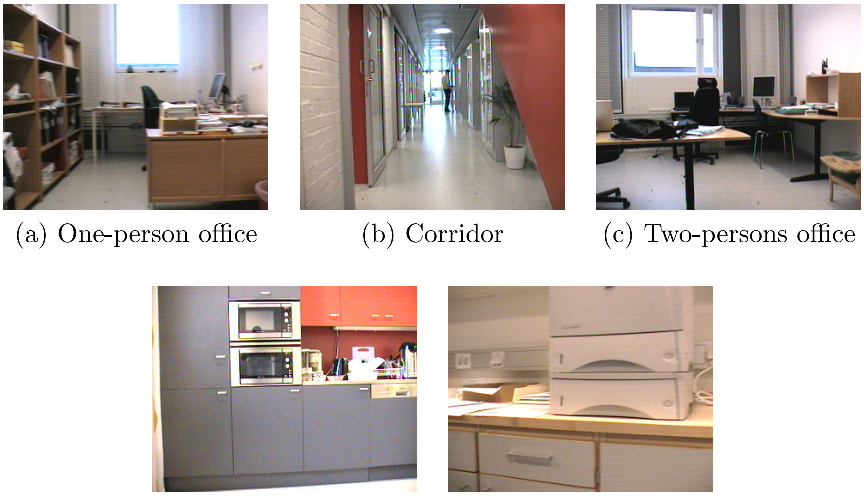
a) The twelve original sequences taken from the KTH-IDOL1 database, which can be seen as an extension of the KTH-INDECS database [1] on mobile robot platforms. The sequences were acquired under various illumination conditions over a span of two weeks capturing natural variability of an office environment. The KTH-IDOL1 database was first presented and used in [2] for evaluating the robustness of a visual place recognition system.

b) Additional sequences acquired six months later using the same experimental scenario. The motivation for this extension was to obtain data for experiments testing the ability of our algorithms to adopt to a wider spectrum of variability that occurs within dynamic environments over a long period of time.

The whole database is freely available on the Internet and can be obtained at `http://cogvis.nada.kth.se/IDOL/`.

The rest of the report is organized as follows: Section 2 provides information about the office environment in which the database was acquired. Section 3 introduces the employed robot platforms, and Section 4 describes the acquisition procedure. Then, Sections 5 and 6 present interesting examples illustrating attributes of the database and captured variability. Finally, Section 7 discusses 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 Environment



(a) One-person office (b) Corridor (c) Two-persons office

(d) Kitchen (e) Printer area

Figure 1: Exemplary images taken from one of the sequences presenting the interior of each of the five rooms.

All the image sequences included in the database were acquired in The Computational Vision and Active Perception Laboratory, at the Royal Institute of Technology, Sweden. The part of the environment under consideration consists of five rooms with different functionality located on the same floor: *the corridor*, *the printer area*, *the kitchen* , a *two-persons office* and a *one-person office*. Exemplary pictures captured in each of the rooms are shown in Figure 1. A general map of the environment is presented in Figure 2. These rooms are physically separated by sliding glass doors. However, there is one exception. *The printer area* is in fact a continuation of *the corridor*, and was treated as a separate room due to its different functionality. Therefore, the boundary between these two rooms can be regarded as arbitrarily marked.

As it was stated previously, the rooms fulfill different functions which determines the furniture layout, objects inside as well as the activity that is likely to occur. Places like *the corridor*, *the printer area*, and *the kitchen* can be regarded as public, which implies that various people may be present. On the other hand, the offices were imaged when they are empty or with their owners at work. In *the corridor* and *the printer area*, furniture is mostly fixed and objects are less moveable. As a result, these areas were less susceptible to variations compared to *the kitchen* and *the offices* in which the furniture (e.g. chairs) is moved more often, objects (e.g. cups) were taken in/out, and decoration was also changed by the owner. Moreover, the whole environment, especially the regions close to windows, is heavily influenced by external illumination and weather conditions. For instance, in sunny weather, the shadows and reflections caused by sunlight will affect the visual appearance of the rooms.

It is also worth pointing out that the database was recorded only in part of the laboratory. However, because of the transparent glass door, some additional offices could still be visible in the images taken in the corridor.

3 Robot Platforms

The images sequences, laser scans and odometry data were acquired using two mobile robot platforms, the PeopleBot Minnie and the PowerBot Dumbo (see Figure 3). The robots were controlled manually using a wireless joystick. Both robots were equipped with a perspective camera, a SICK laser scanner and wheel encoders.

3.1 Camera

The images were taken using a pan-tilt-zoom Cannon VC-C4 camera mounted on the robots, and acquired at a speed of 5 frame per seconds. As shown on the images, the cameras were mounted at different heights. All the images were acquired using the same camera settings. Detailed parameters and settings of the cameras on both robots can be found in Table 1.

Platform	Minnie	Dumbo
Frame Rate	5fps	
Resolution[2]	320×240	
Exposure	Auto	
Focus	Auto	
Filed Angle	Horizontal:45° Vertical:35°	
Height	98cm	36cm
Tilt[3]	0°	13° (Up)

Table 1: Parameters and settings of the cameras.

3.2 Estimation of the Pose of the robots

For the purpose of labeling, the position and orientation, i.e. the pose, of the robot was tracked using a laser-based localization method [3]. A description of the file format used to store laser scans and odometry data can be found in Appendix B.

4 Data Acquisition

Each image sequence was continuously acquired at a rate of 5-fps using the perspective camera, while the robot was manually driven (average speed around 0.3-0.35m/s) along a roughly planned path (see Figure 4) in the lab. The path was designed in a way allowing the robots to capture the visual appearance of all the rooms.

As it was already mentioned, similar path was followed during the acquisition of all the sequences. However, due to manual control, obvious differences in viewpoints still existed between different sequences. To illustrate the differences, a comparison of two paths is presented in Figure 5a. Additionally two sets of

[1]All the images in the database were cropped into the size 309×240 because the Cannon VC-C4 Camera always produces ten pixel width black line on the left margin of the picture.

[2]The camera on Dumbo was tilted up in order to reduce the amount of floor captured in the images

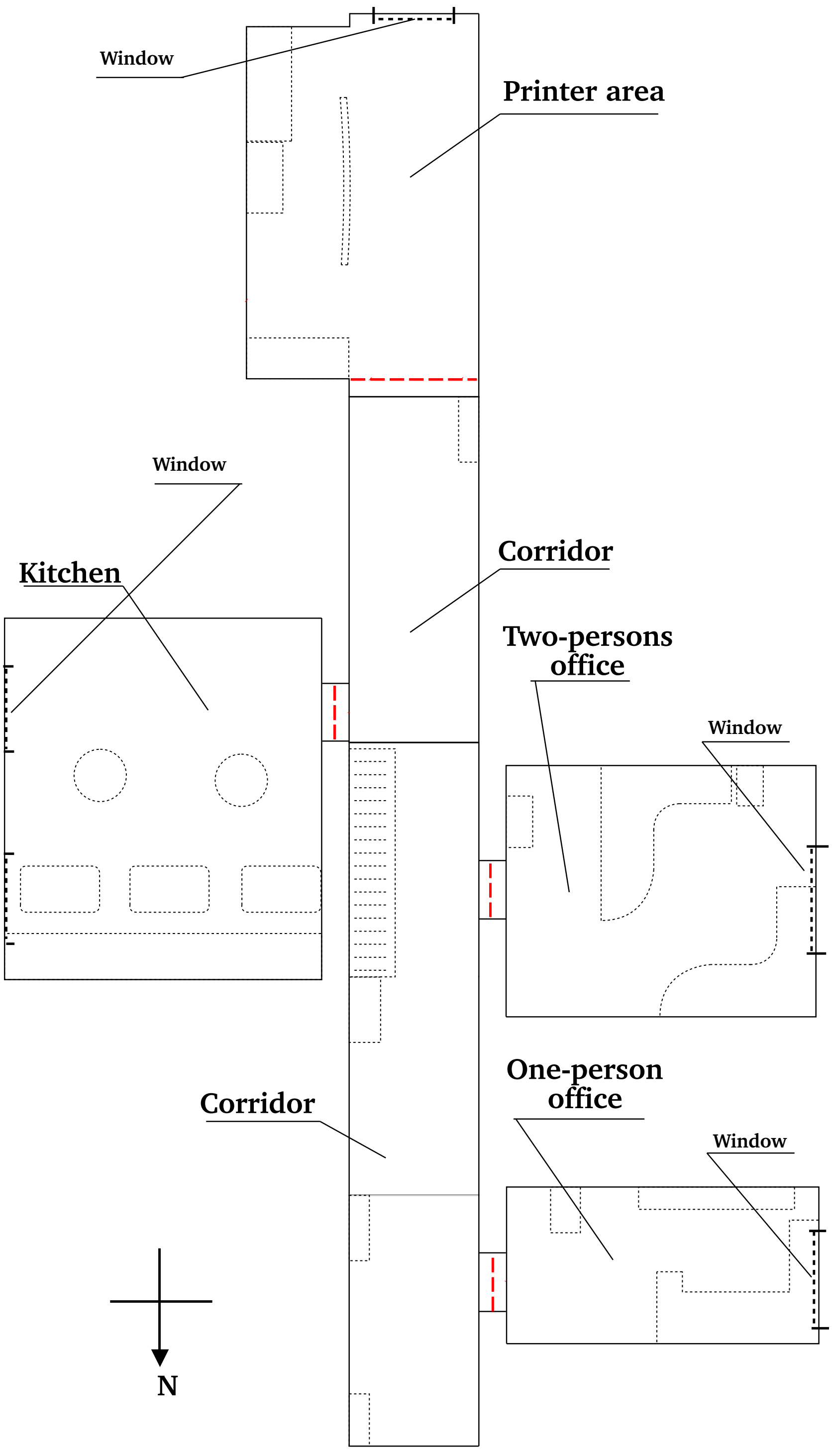
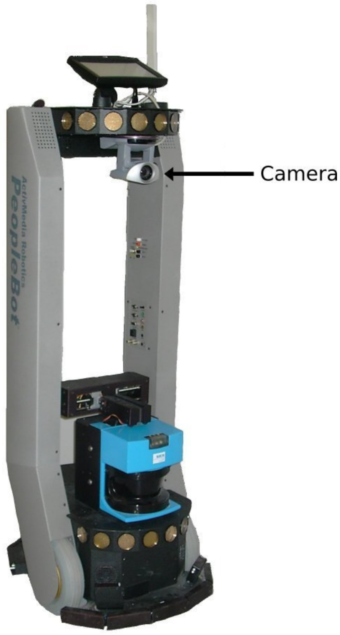


Figure 2: A general map of the environment. Boundaries between the five rooms were marked with dashed lines. Outline of large pieces of furniture and locations of windows were also marked on the map with dotted lines. Small and portable pieces of furniture such as chairs were not marked on the map.



(a) Minnie

(b) Dumbo

Figure 3: The two robot platforms used for data acquisition.

images from two sequences, recorded at the closest points, are shown in Figure 5b. Plots showing acquisition paths for all the sequences are attached to the database.

The artificial light was always on in all rooms during acquisitions. However, it was not fixed or specially adjusted before the acquisition process, which meant the light condition was the same as everyday working environment, and it would not eliminate the influence of external illumination and weather changes.

Four sequences were recorded, for each robot platform (Minnie and Dumbo) and for each weather condition (sunny, cloudy, and night). Of these four sequences, the first two (inherited from KTH-IDOL1) were acquired six months before the last two. Therefore, 24 image sequences were recorded in total.

An image sequence usually contains 800 − 1100 frames. Each image was automatically labeled with a timestamp, robot's pose, and the room in which it was acquired. The labeling was based on the position and direction of the camera given by the aforementioned laser-based localization system. This pose was represented in a defined global coordinate system (Appendix C). Then, the estimated coordinates were used to determine in which room the image was acquired.

5 Captured Variability

The image sequences were acquired in real-world setting, over a span of 6 months, and under different illumination and weather conditions. Consequently, different visual variations in a indoor environment were captured in the sequences. These variations could be categorized into following categories.

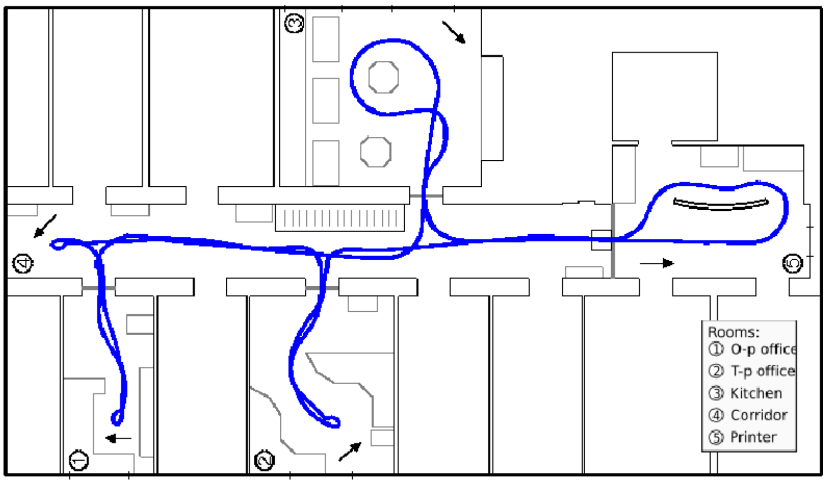


Figure 4: Path of the robot doing acquisition of one of the sequences. The arrow indicates the direction of driving and the start point is marked with square box.

5.1 Variations Introduced by Illumination Change

Very significant changes were caused by the illumination. To capture this variability and intrinsic properties of the environment, sequences were acquired under three illumination and weather conditions: in *cloudy* weather, in *sunny* weather, and at *night*. The influence could be easily observed from the exemplary images presented in Figure 6. Especially, because of the unpredictability of the sunlight, the shadows and reflections caused by the sunlight heavily influenced the visual appearance of the captured scenes. And due to the fact that auto-exposure model of the cameras were turned on, the images became dark when acquired in front of the window in sunny days.

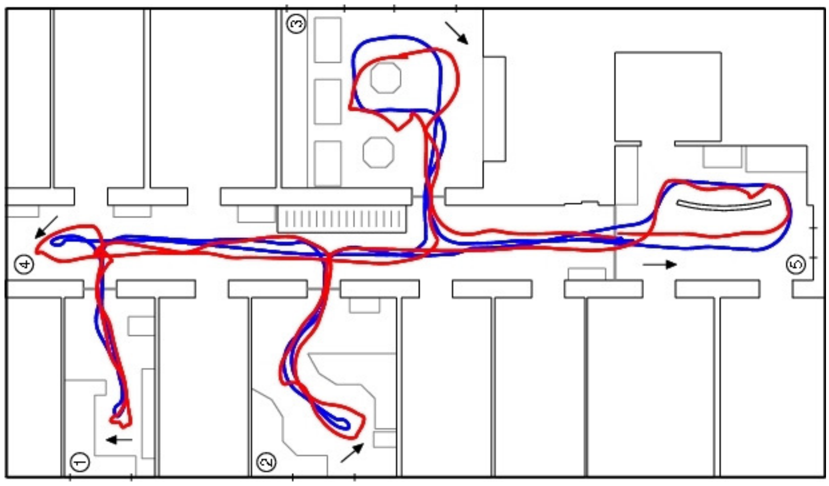
5.2 Environment Variations Caused by Human Activity

The visual appearance of the indoor environment was changing over time since it was being used. The following variations introduced by human activity can be observed in the images:

- People appeared in different rooms during working time (Figure 7a).
- Objects were moved and taken in/out of the drawers (Figure 7b).
- Pieces of furniture, such as chairs, were pushed around (Figure 7c).
- Decoration of one of the rooms was changed by the owner: new articles and painting were added/removed, computers and chairs were replaced (Figure 7d).

6 Difficult Examples in The Database

During the acquisition, we encountered several problems which may also occur in a practical applications. The images in the database were labeled according to the position of the robot at the moment of acquisition. As a result, when the robot faced a transparent glass door of one of the offices or was positioned in the transition region between two rooms, the visual information



(a) The difference between robot path for two acquisitions.

(b) Images illustrating the difference of viewpoints for images acquired closest to each other taken from two different sequences.

Figure 5: Differences caused by manual control of the robot.

captured in the image was not consistent with its label. Moreover, during the acquisition, the robots also acquired images which contained very little visual cues (e.g. images of blank walls).

Another difficulty was caused by the motion of the robots. Part of the images were blurred due to the fact that the robot was shaking while driving and autofocus function of the camera could not react fast enough during rotations. This makes the database suitable for testing of applications that should operate on a mobile platform where such problems are common.

7 Future Extension to the Database

Several enhancements of the database are planned in the future. Currently, the database contains images acquired on one floor of the laboratory. We plan to reacquire the database in other office environments using the same strategy. It would also be of interest to add more rooms to the database and extend it again after some time. As mentioned before in Section 6, the visual information captured in some images is inconsistent with the label. Using an omnidirectional camera together with standard perspective camera could be used to solve this ambiguity.

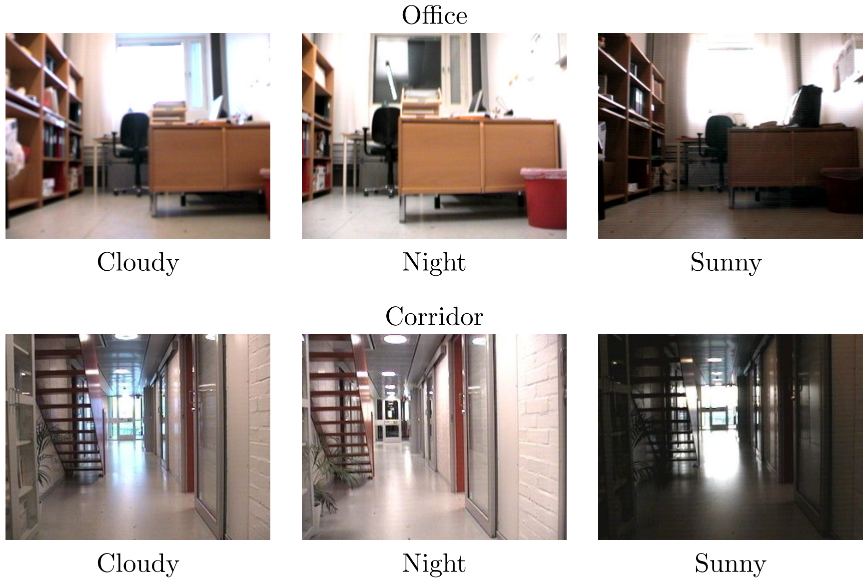


Figure 6: Example images acquired under different illumination conditions

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

(b) Objects were moved and new objects appeared.

(c) Pieces of furniture were moved around.

(d) Decoration was changed.

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

References

[1] A. Pronobis, B. Caputo. The KTH-INDECS database. Technical Report 297, KTH, CVAP, 2005. Available at http://cogvis.nada.kth.se/INDECS/

[2] A. Pronobis, B. Caputo, P. Jensfelt, H. I. Christensen. A discriminative approach to robust visual place recognition. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems* (IROS'06), Beijing, China, Oct. 2006.

[3] J. Folkesson, P. Jensfelt, and H. Christensen, Vision SLAM in the measurement subspace. in *Proceedings of the IEEE International Conference on Robotics and Automation* (ICRA'05), Barcelona, Spain, Apr. 2005.

A File Naming Conventions

Complete information about each picture 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*⟩_r⟨*room*⟩_x⟨*x.coordinate*⟩_y⟨*y.coordinate*⟩_a⟨*angle*⟩.⟨*extension*⟩

⟨*timestamp*⟩ - The internal time of the robot when the image was taken.

⟨*room*⟩ - Code of the room where the picture was taken.

::=	BO	- One-person office
|	CR	- Corridor
|	EO	- Two-persons office
|	KT	- Kitchen
|	PA	- Printer Area

⟨*x.coordinate*⟩ - X-Coordinate in the global coordinate system (unit:m).

⟨*y.coordinate*⟩ - Y-Coordinate in the global coordinate system (unit:m).

⟨*angle*⟩ - Angle representing the direction of the camera (unit:*rad*).

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

The internal clocks of the robot were used in order to timestamp the images. For some sequences, there might be a few hours deviation with respect to the real time. Detail information about the global coordinate system can be found in Appendix C.

B Odometry and Scans File Format

The odometry and laser scanner data collected during the acquisition were provided together with the sequences. The odom and scans files[1] were written in a format using in the "cure/toolbox" software package[2], which is a C++ based software library providing utility algorithms for robot control. Each row in the files contained one object, a pose in the odometry file or a laser scan in the scan file. The line format of files for both types is 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

Values in the files are meaningless, or just used for bookkeeping in order to keep track of how many data that have been written etc. These bookkeeping value are represent by "Ø" above. For odometry file, only x,y,theta and the time are used and z,phi,psi will always be zero.

[1]Names of the odometry and laser scanner data file.
[2]http://www.CURE.org

C Coordinate System Used In Labeling

The estimated pose of the robot is labeled using a global coordinate system. Figure 8 presents the map of the environment within the coordinate system. The positive X-axis defines the angle zero of the direction of the camera, and the angle value becomes positive when moving toward the positive Y-axis. While labeling, the coordinate have been shifted from the center pose of the robots to the position of camera. However, for simplicity the same offset was used for Minnie and Dumbo.

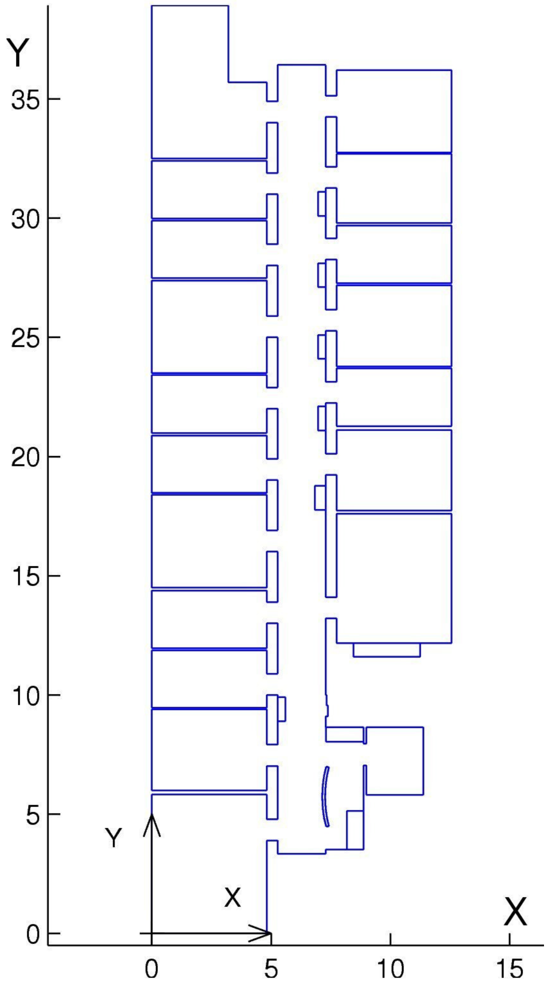


Figure 8: Coordinate system used for position estimation

D Structure of IDOL2 Database

This sections will give information about the directory structure of the database. For sake of simplicity, the database is divided into subdirectories according to image sequences. The full directory structure is presented as follow:

```
⟨root⟩
|
|
+—<robot>                – Robot platform
     |
     |
     +–<sequence_id>     – Sequence ID
         |
         |
         +–images\       – Directory containing images
         |
         |
         +–movie.mov     – Movie created from the image sequence
         |
         |
         +–path.jpg      – Map with plotted acquisition path
         |
         |
         +–odom.dat      – Odometry data file
         |
         |
         +–scans.dat     – Laser scans data file
```

⟨*robot*⟩::=dumbo|minnie

⟨*sequence_id*⟩::=<illumination><no>
    ⟨*illumination*⟩::=cloudy|night|sunny
    ⟨*no*⟩::1|2|3|4