Learning to Open New Doors

Ellen Klingbeil, Ashutosh Saxena, Andrew Y. Ng
Computer Science Department,
Stanford University, Stanford, CA 94305
{ellenrk,asaxena,ang}cs.stanford.edu

Abstract

As robots enter novel, uncertain home and office environments, they are able to navigate these environments successfully. However, to be practically deployed, robots should be able to manipulate their environment to gain access to new spaces, such as by opening a door and operating an elevator. This remains a challenging problem because a robot will encounter doors it has never seen before.

Objects such as door handles and elevator buttons, though very different in appearance, are functionally similar. Thus, they share some common features in the way they can be perceived and acted upon. We present a vision-based learning algorithm that captures these features to: (a) find where the door handle is located, and (b) infer how to manipulate it to open the door. Our system assumes no prior knowledge of the 3-D location or shape of the door handle. We also experimentally verify our algorithms on doors not seen in the training set, advancing our work towards being the first to enable our robot to navigate anywhere in a new building by opening doors and elevators, even ones it has not seen before.

Introduction

There has been recent interest in using robots not only in controlled factory environments but also in unstructured home and office environments. In the past, successful navigation algorithms have been developed for robots in these environments; but to be practically deployed, robots must also be able to manipulate their environment to gain access to new spaces, such as by opening a door and by operating an elevator. This remains a challenging problem because a robot will likely encounter doors and elevators it has never seen before.

In robotic manipulation, most work has focused on developing control actions for different tasks, such as grasping objects (Bicchi & Kumar 2000), assuming a detailed 3-D model of the environment is known. There has been some recent work in opening doors using manipulators (Rhee *et al.* 2004; Petersson, Austin, & Kragic 2000; Kim *et al.* 2004; Prats, Sanz, & del Pobil 2007); however, it was focused on developing control actions assuming a known location of a known door handle. (Petrovskaya & Ng 2007) assumed a known detailed model of the door and door handle to be opened. In practice, a robot has to rely on only its sensors to be able to perform manipulation in a new environment, and current sensor technology does not have enough resolution to build a detailed model of the object that is required for manipulation purposes.

Most work in computer vision has focused on object recognition, e.g. (Serre, Wolf, & Poggio 2005). However, for manipulation purposes, a robot not only needs to locate the object, but also needs to find out what to do with the object. For example, if the intention of the robot is to enter a door, it needs to find out where the door handle is as well as determine what action it must take in that situation—turn the door handle right and push for example.

Our work does not assume existence of a known model of the object (such as the door, the door handle, or the elevator button). Instead, we focus our work on the problem of manipulation in novel environments, in which a model of the objects is not available. We also demonstrate the robustness of our algorithms through extensive experiments in which the robot was able to reliably open new doors in new buildings, even ones which were seen for the first time by the robot (and the researchers working on the algorithm).

Algorithm Overview

Our perception system consists of two parts: (a) Object Perception: finding the object, and (b) inferring how to manipulate the object.

For finding the object, we compute features that were motivated in part by some recent work in object recognition (Serre, Wolf, & Poggio 2005) and robotic grasping (Saxena *et al.* 2006). We use the Support Vector Machines (SVM) (Vapnik 1995) learning algorithm and select the most relevant directions using Principal Component Analysis. We also take advantage of some contextual information to learn a location based prior (partly motivated by (Torralba 2003)). This captures properties such as that a door handle is less likely to be found close to the floor. To deal with multiple handles/buttons in an image and the spatial correlation between their predicted locations (see Figure 1), we used a K-means clustering algorithm to return the center of each handle.

We estimate the 3-D location of the handle/button from the 2-D location in the camera frame and from a horizontal laser scan, by assuming that the walls are vertical to the ground.

Given a rectangular image patch containing an object, we then need to classify what action to take. We consider three types of *abstract* actions: turn left, turn right and press. To distinguish between such actions, we used a similar classifier (as described above) and achieved an overall classification accuracy of 94.1%.

With the 3-D location and desired abstract action type known, we now define each abstract action as a set of key-

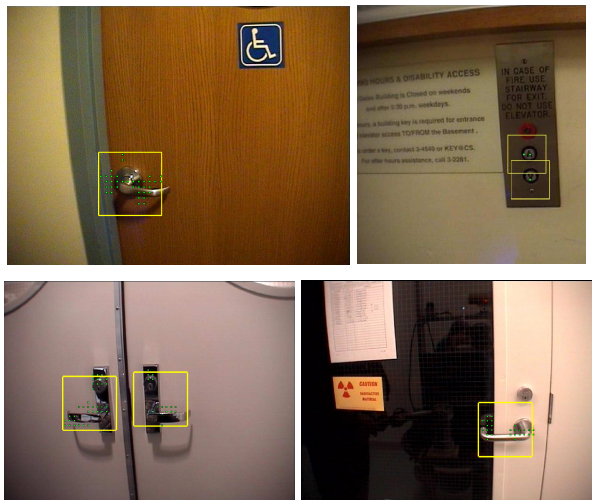

Figure 1: Some typical results showing the door handles and the elevator buttons found.

points, through which the robot (and its arm) has to pass. We used a motion planning algorithm for inferring these keypoints. (Schwarzer, Saha, & Latombe 2005)

Experiments

Our robotic platform consists of a 5-dof position-controlled robotic arm with a parallel plate gripper (Katana, by Neuronics) mounted on a Segway platform. Our vision system uses a pan-tilt-zoom (Sony DV100) camera and a 2-D laser scanner (Hokuyo) mounted on a frame behind the robotic arm.

An experiment began with the robot started at a random location within 3m of the door. It used lasers to navigate to the door, and our vision-based classifiers to find the handle.

In the experiments, our robot was seeing all of our test locations for the first time. The training images for our vision-based learning algorithm were collected in completely different buildings, with different doors, structure, ambient decoration, etc. We tested our algorithm on two different buildings on a total of five different floors (about 20 different types of doors). For each door, test cases were also run where the robot localized at different angles, typically between -30 and +30 degrees with respect to the door, to verify the robustness of the algorithm.

We achieved an average recognition accuracy of 94.1% and a classification accuracy of 97.1%, leading to a success rate of 91.2% in a total of 34 experiments (see Table 1).[1] Notable failures among the test cases included glass doors (erroneous laser readings), doors with numeric keypads (classification error due to confusion with elevator buttons), and very dim/poor lighting conditions. Due to the small size of the elevator buttons (1 inch diam) and that the arm-vision system was calibrated only up to an accuracy of 2 cm, they were more difficult to push reliably.

Videos of the robot opening new doors and elevators are available at:
http://stair.stanford.edu/multimedia.php

[1]The localization error is the mean error (in cm) between the predicted and actual location of the door handle.

Table 1: Error rates obtained for the robot opening the door in a total number of 34 trials.

Door Type	Num of trials	Recog. (%)	Class. (%)	Localization (cm)	Success-rate
Left	19	89.5%	94.7%	2.3	84.2%
Right	15	100%	100%	2.0	100%
Total	34	94.1%	97.1%	2.2	91.2%

In conclusion, we have developed robust algorithms to significantly advance our work towards being the first to enable our robot to navigate anywhere in a new building by opening doors and elevators, even ones it has not seen before.

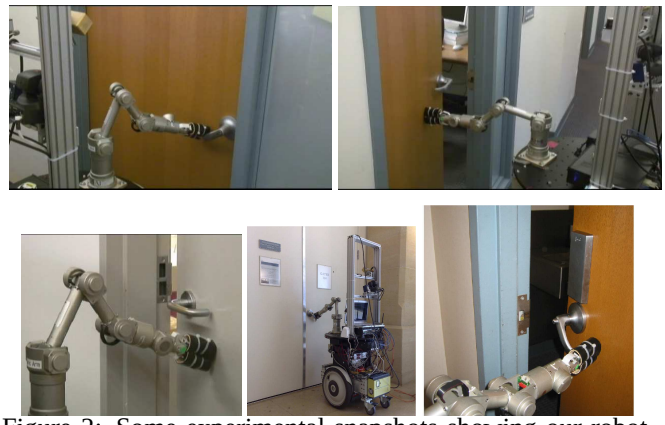

Figure 2: Some experimental snapshots showing our robot opening different types of doors.

Acknowledgments We thank Andrei Iancu, Srinivasa Rangan and Morgan Quigley for their help in the experiments.

References

Bicchi, A., and Kumar, V. 2000. Robotic grasping and contact: a review. In *ICRA*.

Kim, D.; Kang, J.-H.; Hwang, C.; and Park, G.-T. 2004. Mobile robot for door opening in a house. *LNAI* 3215:596–602.

Petersson, L.; Austin, D.; and Kragic, D. 2000. High-level control of a mobile manipulator for door opening. In *IROS*.

Petrovskaya, A., and Ng, A. Y. 2007. Probabilistic mobile manipulation in dynamic environments, with application to opening doors. In *IJCAI*.

Prats, M.; Sanz, P.; and del Pobil, A. P. 2007. Task planning for intelligent robot manipulation. In *IASTED Artificial Intel App*.

Rhee, C.; Chung, W.; Kim, M.; Shim, Y.; and Lee, H. 2004. Door opening control using the multi-fingered robotic hand for the indoor service robot. In *ICRA*.

Saxena, A.; Driemeyer, J.; Kearns, J.; and Ng, A. Y. 2006. Robotic grasping of novel objects. In *NIPS*.

Schwarzer, F.; Saha, M.; and Latombe, J.-C. 2005. Adaptive dynamic collision checking for single and multiple articulated robots in complex environments.

Serre, T.; Wolf, L.; and Poggio, T. 2005. Object recognition with features inspired by visual cortex. In *CVPR*.

Torralba, A. 2003. Contextual priming for object detection. *IJCV* 53.

Vapnik, V. N. 1995. *The Nature of Statistical Learning Theory*. Springer.