User Observation & Dataset Collection for Robot Training

Caroline Pantofaru
Willow Garage, Inc.
Menlo Park, CA 94025
pantofaru@willowgarage.com

Categories and Subject Descriptors: I.5.2 [Computing Methodologies]: Pattern Recognition - Design Methodology, H.1.2 [Information Systems]: Models and Principles - User/Machine Systems

General Terms: Measurement

1. INTRODUCTION

Personal robots operate in human environments such as homes and offices, co-habiting with people. To effectively train robot algorithms for such scenarios, a large amount of training data containing both people and the environment is required. Collecting such data involves taking a robot into new environments, observing and interacting with people. So far, best practices for robot data collection have been undefined. Fortunately, the human-robot interaction community has conducted field studies whose methodology can serve as a model. In this paper, we draw parallels between field study observation and the data collection process, suggesting that best practices may be transferable. As a use case, we present a robot sensor dataset for training and testing algorithms for person detection in indoor environments.

2. THE DATASET

Detecting people in imagery has long been a pursuit of the computer vision community, however the focus has been on single images and video [?]. More recently, automobile automation has spurred research into vehicle-based perception, from either car-mounted video cameras or laser scanners [?]. Only very recently, however, has a focus on perceiving people indoors emerged from the growing personal robotics research community. With the increased availability of personal robot research platforms, it is now possible to collect realistic robot sensor data of people and their environments. The **Moving People, Moving Platform Dataset (MPMP Dataset)** [?] is such a dataset.

Contents: The MPMP Dataset was collected using the Willow Garage Inc. PR2 platform (Fig. **??**(a)) and contains data from multiple robot sensors of people in office environments. The PR2 stands 1.4m tall, making its presence obvious. Onboard sensors included two pairs of stereo video cameras located 1.3m off the ground, a tilting 2D laser scanner 1.2m off the ground, a stationary 2D laser scanner 30cm off the ground, and odometry. Figs. **??**(a,b) show a data snapshot. Approximately two hours of data were collected at four different companies. Ground truth annotations of people's locations were generated for 37,710 out of the 108,124 frames. See the dataset webpage [?] for more information.

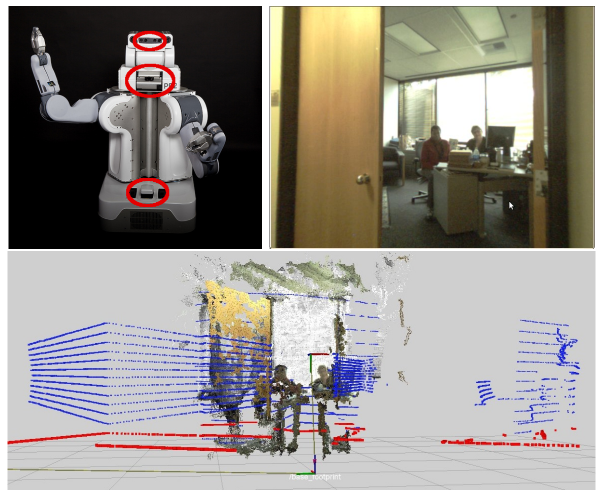

Figure 1: The PR2 robot with sensors used for data collection circled (top left); Image from the MPMP Dataset (top right); Corresponding laser and stereo data snapshot (bottom). (Best viewed in color.)

Dataset Requirements: For the personal robotics scenario, the dataset had to be collected indoors, from a realistic robot, in close proximity to people going about their daily office routine. The perception task required a large quantity of data from a variety of sensors onboard the robot. Also required were sufficiently many scenes containing people (of various numbers), scenes devoid of people, people in different poses (standing, sitting, interacting, etc.), and different backgrounds (a variety of offices, hallways, meeting rooms, etc.). This, in turn meant interacting closely with the subjects, without disturbing or annoying them.

3. USER STUDIES & DATA COLLECTION

A number of common scenarios from user observation play out during data collection. During collection, the robot is the observer, however it does not have the ability to judge which information is important. A human operator provides this higher-level reasoning, making the collection process a fusion between the observations of the robot and operator.

Briefing the Audience: As discussed in [?], the first step of an interview is an introduction to the data collection

HRI'11, March 6–9, 2011, Lausanne, Switzerland.
ACM 978-1-4503-0561-7/11/03.

personnel, process and purpose. This includes discussing public parts of data, and obtaining permission to record. For this, we used a formal presentation and demonstration at the collection site, allowing participants to physically interact with and tele-operate the robot.

When collecting a public dataset, it is important to allow subjects to opt out. Subjects could view the data and request that sections be removed. Over four collection sessions with 10 to 40 subjects, three implicitly opted out by closing office doors, and one opted out after reviewing the data.

Note two key differences between training datasets and user study data: (1) the data is public and can be seen by the subjects, and (2) removing a small amount of data does not significantly alter the quality of the dataset, and more data can always be gathered to rectify bias.

Real Environments Provide Context: There are two location options when observing users; bring them in-house or go to them. Bringing users in-house is logistically easier, set-up of the test bed is only done once, and extraneous variables such as interruptions can be controlled. However, going into a user's environment provides more realistic data including environmental factors, normal work practices, and comfort levels. The data collected in a user's environment are ongoing and concrete, not a summary or abstraction [?].

A parallel situation exists for data collection. Many datasets are collected in a lab (e.g. [?]), controlling the environment, subjects' clothing and activities; simplifying the data and its collection. On the other hand, collecting the MPMP Dataset at real venues (offices) provided context for subjects' appearance and actions. Subjects dressed appropriately for work, acted more naturally, and performed tasks that were necessary to their workday. In addition, obtaining realistic background and lighting examples is extremely important for the performance of many computer vision algorithms.

Autonomy versus Tele-operation: In the ideal data collection scenario, the robot behaves as during normal operation and people are observed reacting. For the MPMP Dataset, this would involve PR2 navigating autonomously, pursuing office-related activities such as delivering objects, interacting with people in a task-appropriate manner, all while collecting perception data. However, this is not always possible, due to lack of current functionality, or the difficulty in ensuring the robot stays within a prescribed area and obeys (often subtle) rules of office etiquette.

For such reasons, there are circumstances in which some Wizard-of-Oz or tele-operation is preferable even if autonomy is possible or more realistic. For the MPMP Dataset, an operator used a joystick to drive the robot around, directing it to acceptable locations, and controlled when data were recorded. This had the advantage that recording could be paused if a person who had opted out of the dataset was encountered, if anyone in the environment was acting in an 'artificial' manner (such as taking photographs), or if the environment the robot was driving through was particularly monotonous or repetitive (such as an empty hallway).

Tele-operation did bias the data, however. The operator followed social conventions such as staying on one side of a hallway, whereas autonomous navigation would have followed the safest route down the center. By controlling recording, the operator biased the type and amount of data. Most importantly, the presence of the operator changed people's behavior, passersby felt compelled to acknowledge the operator and stop to chat. It is possible to tele-operate the PR2 remotely from another room, alleviating some of the subject-operator interaction issues. Unfortunately, the host companies would not give permission to separate the operator from the robot. Also, the wireless network was often too weak for reliable long-range tele-operation.

As in user studies, Wizard-of-Oz or tele-operation is a tool to explore unfinished systems or avoid legal issues. However, its influence on the data must be considered and disclosed.

Beyond Novelty, Before Aggravation: The length of time that a robot spends in a human environment can affect the behavior of its inhabitants. At first, it is a novelty, causing unusual behavior (such as photograph-taking) that is not representative of long-term interactions with the robot. One benefit of novelty during data collection, however, was that subjects were patient with the robot invading their workspace.

By the end of each day-long collection session, however, the subjects grew less patient. The noise of the PR2 driving was an issue, and constant observation by the robot and operator made people weary. Most of the subjects were polite despite their impatience as they supported the end product.

For a period in the middle of the day, subjects were accustomed to the robot, and their patience was not exhausted. This was the best time for data collection as people acted most naturally, generally ignoring the robot and going about their work. Ideally, this period could be extended. Spreading data collection over multiple days, with shorter sessions, might extend this period, however this was cost-prohibitive and impractical for the MPMP Dataset.

Another possibility would be to have the robot perform a useful task in parallel with data collection. We hypothesize that making the robot more useful would increase subjects' tolerance of its presence. For example, the robot could fetch and deliver objects, collecting data while doing so. The chosen task, however, must not interfere with data collection by biasing the robot's location, people's behavior, or by allowing objects to occlude sensors.

4. CONCLUSION

Personal robots have many things to learn and require a large quantity of data to learn them. Whether learning by demonstration, by trial and error, or collecting datasets for perception, robots will need to collect vast amounts of data without burdening the subjects. The parallels between gathering data for robot training and observing users during studies suggest the application of user study methodology as a basis for data collection methodology. Given the wide array of possible data, robotic platforms and algorithms, it is too early to set strict guidelines on collection practices. A clear set of guidelines, however, on how to report collection methodology and possible biases would benefit the community.

5. REFERENCES

[1] H. Beyer and K. Holtzblatt. *Contextual Design*. Morgan Kaufmann, 1998.

[2] CMU. Motion Capture Database. `mocap.cs.cmu.edu`.

[3] N. Dalal. *Finding People in Images and Videos*. PhD thesis, INPG/INRIA, Grenoble, France, 2006.

[4] P. Dollár, C. Wojek, B. Schiele, and P. Perona. Pedestrian Detection: A Benchmark. In *IEEE Conf. on Computer Vision and Pattern Recognition*, 2009.

[5] C. Pantofaru. The Moving People, Moving Platform Dataset. `http://bags.willowgarage.com/downloads/people_dataset.html`.