Navigating among people in crowded environment: Datasets for localization and human robot interaction

Rafael Ramón-Vigo[a], Javier Pérez-Lara[a], Fernando Caballero[b] and Luis Merino[a]

Abstract— **This paper presents a novel set of data for the evaluation of visual place recognition in both indoors and outdoors environment in addition to sensor information to evaluate human-robot interactions in crowded areas. The datasets were recorded in the Royal Alcázar of Seville (Spain), a tourist hotspot that may have more than 5000 visits per day. We recorded a large set of images sequences from a stereo camera and scan measurements from three laser mounted on a moving robot. The datasets are timestamped and stored by means of the well-known Robot Operating System (ROS) log functionality. The robot traveled more than one kilometer in each experiment, and every trial was performed at different time of the day so we could capture the evolution of lighting conditions over the images. The tourist attendance also depends on the hour, providing datasets with a lot of examples to model into a social-way the different places such as corridors, gates, queues, groups of people, etc.**

I. INTRODUCTION

In this paper we present a large dataset of images covering a public space, the Royal Alcázar of Seville (which may have more than 5000 visits per day). The recorded data encompass different timetables along four days of experiments and is intended for localization estimation in crowded areas and for human robot interaction (HRI) research. This data will be of particular interest for researchers in topics like scan matching, visual localization, people detection and tracking, and socially-spatial modeling.

The datasets presented here are noteworthy for several reasons. First, they include a large collection of outdoors and indoors data taken from a mobile robot. It is composed by 10 complete teleoperated tours of 1.8*km* approximately each, making a total of nearly 20*km* for the entire datasets. An important difference with respect [17] and [15] is that they collected data from fixed zenithal cameras. Furthermore, this paper presents a set of experiments recorded at different times, providing an important database of images of the same places but with different light and crowded conditions. This represents a complete basis of evaluation for several algorithms and techniques that sustain themselves in image processing of the environment.

Many datasets are publicly available for testing localization algorithms, like New College Dataset [19], a single 2.2*km* trajectory where a Bumblebee is used as stereo pair in addition to two working lasers in orthogonal planes, or St. Lucia Suburbs dataset [9], a single journey through the suburb of St Lucia, Queensland, Australia, where the large scale route was traversed five times during the same day to gather 66*km* of data from webcam, but no lasers were used. Both outdoor datasets are used to research on loop-closure detection or visual odometry [10], [8] by taking GPS as ground truth pose. In [4] is presented a bank of 6 outdoor datasets also based on visual odometry and GPS in a trajectory of 6*km* outside of the School of Engineering at the University of Málaga, mainly road and parking during two different days.

Our contribution focuses on autonomous mobile robots at pedestrian level, in a crowded GPS denied area, where reliability, accuracy and human avoidance becomes essential. In the presented work there are enough datasets covering nearly full daylight scenario variations that can be used for robot training and the rest for algorithm validation. Each experiment has been intentionally carried out following the same trajectory approximately, starting at the same point, making the robot trajectories in a similar way and also finishing at the same point. All these elements configure an important basis to evaluate distinct techniques over the several topics that we mentioned before, with a practical set of complementary data for 24/7 localization and navigation analysis in human environments.

The presented datasets are also novel from the point of view of interaction between robot and people surrounding. In contrast to [21], these datasets gather the reaction of people at several places such corridors, gates, open areas, etc. and all the data is collected by the onboard robot's sensors. All the people reactions captured are directly motivated by the presence of the robot, so techniques like the presented at [12] could be applied to these datasets instead of trying to apply from observations between people interaction. In addition, we provide a ground-truth for the robot position with an approximate accuracy of 20*cm* in those narrow areas and a 40 *cm* of accuracy in wide areas where the lasers do not reach significant parts of the map.

The paper is structured as follows: Section II describes the datasets and the experiments carried out with the robot. Sections III and IV summarize the interest of the datasets for human-robot interaction and visual place recognition purposes. Finally, Section V presents the conclusions and the future work.

This work is partially funded by the European Commission 7th Framework Programme under grant agreement no. 288235 (FROG) and the project PAIS-MultiRobot, funded by the Junta de Andalucía (TIC-7390)

[a]Rafael Ramón-Vigo, Javier Pérez-Lara and Luis Merino are with the Pablo de Olavide University, Seville (Spain) {rramvig,jiperlar,lmercab}@upo.es

[b]Fernando Caballero is with University of Seville, Seville (Spain) fcaballero@us.es

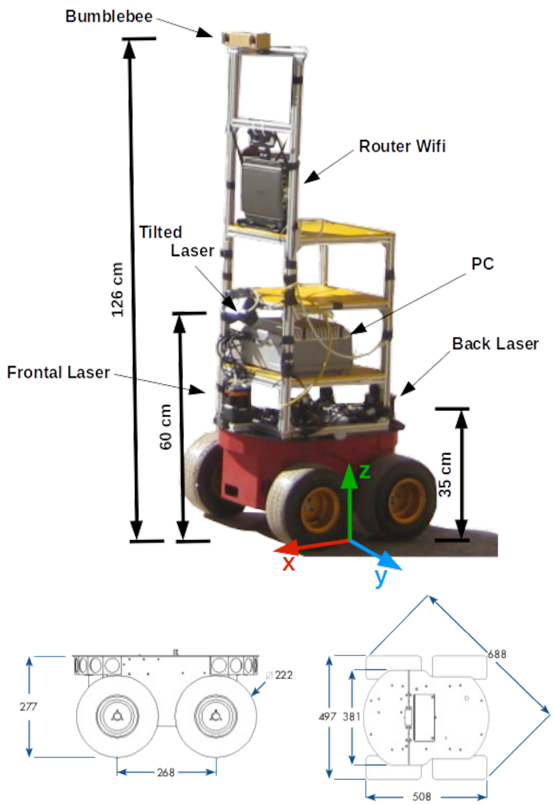


Fig. 1. Top: General platform configuration: Location of the lasers and the bumblebee camera, plus others components. The origin frame is also indicated. Bottom: Dimensions of the platform. Measurements are in *mm*. Source: http://www.mobilerobots.com/ResearchRobots/P3AT.aspx

II. DATA DESCRIPTION

A. Robot Platform

The robot platform used for the datasets is a Pioneer 3AT with a simple aluminum structure to place the sensors and the computer. The sensors and the position into the robot can be seen in Fig. 1, they are the following:

- A stereo camera facing forward at 1.2*m* height.
- Two Hokuyo UTM-30LX placed parallel to the floor facing forward and backwards.
- A Hokuyo URG-04LX tilted 30° in front of the robot.
- Encoders in the robot base for odometry computation.

As commented before, ROS[1] was used as main development tool for data gathering and logging, so all the sensors are recorded using the standard interfaces of ROS and also its communication facilities (topics, messages, filters, etc). Thus, the relative position of all sensors with respect the robot base are encoded using ROS Transforms (TF)[2]. All the sensors are time stamped into the same computer, so they are synchronized.

B. Experiments overview

During each experiment the robot was manually guided through the Royal Alcázar, following almost the same trajectory with small differences between experiments. The trajectory is 1.8*km* long approximately, and the robot passes through large rooms, corridors, small patios and big squares. Figure 2 shows some images of the environment during the robot's translation. The robot always starts in the same position and also finishes in that position, making easier initialization and analysis of data collected.

In order to validate place recognition algorithms and also to localize elements into a single reference frame, a map of the Royal Alcázar was built using a different dataset; this map is also provided with the dataset. Thus, a ground-truth has been computed based on this map, rangefinders and Monte Carlo Localization algorithm. Our visual tests show that the robot is always well localized into the map with errors from 20*cm* (most of the time) to 40*cm* (in large open areas scan matching may have poor likelihood). Although we cannot compare with respect a localization ground-truth, both trajectory and localization into the map are coherent with the real robot motion.

On the other hand, images and lasers can be used to detect and track persons around the robot. This information is valuable in order to learn and model how people interact with mobile robots. While we are not providing yet the person annotation in images and lasers, there already exists algorithms implemented in ROS that automatically detect persons using laser segments classifiers [1] or Histogram of Oriented Gradients (HOG) in images [6].

The experiments were carried out during four consecutive days. We performed 10 different experiments at different times of the day from 8 am to 6 pm (see Table I for details). The experiments are not in a single day because of the needed time to recharge batteries between experiments. Although the ideal setup would be to have all the datasets in a row, we certified that neither the sunlight nor the tourist attendance changed significantly along the four experiments days, so it may not affect the datasets quality.

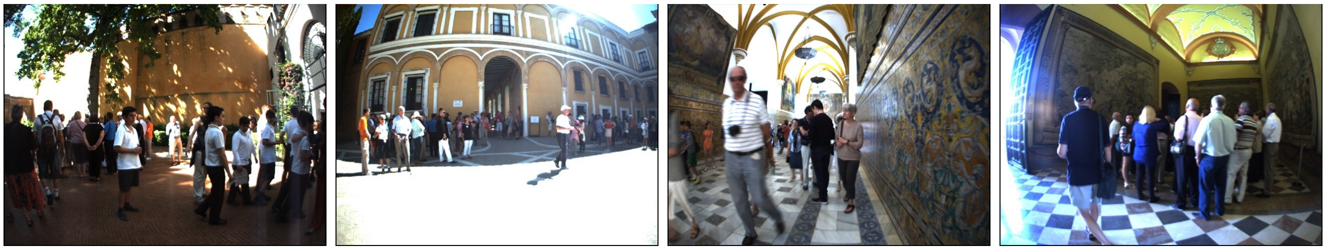
C. Dataset contents

The datasets presented in this paper are available at: `www.upo.es/isa/lmercab/datasets`. Each of the 10 datasets (see Table I) are stored separately in three files: one for sensors measurements, other for raw RGB images and a third one with the rectified grayscale images. All the datasets are logged and also processed using ROS tools, such as the *ROS Bag*[3]. The different information stored into the logs also follows ROS interfaces and development main guidelines, so that the reader can understand easily the dataset with minimum ROS background.

The three log files per dataset contain the following information:

- Robot odometry. Topic */pose*, datatype *nav_msgs/Odometry*[4]. This odometry corresponds to a Pioneer 3AT robot. It is a differential platform and the odometry is computed by incremental distance

[1] http://wiki.ros.org/
[2] http://wiki.ros.org/tf
[3] http://wiki.ros.org/Bags
[4] http://wiki.ros.org/nav_msgs



(a) (b) (c) (d)

Fig. 2. Images gathered by the robot while teleoperated at Royal Alcázar. From left to right: Lion Gate Courtyard, Hunting Courtyard, Vault Room and Tapestry Room.

traveled measurements from the wheel encoders. The information is published at a rate of 10 *Hz*.

- Laser measurements. The robot is equipped with 3 Hokuyo Lasers, one frontal (topic */scanfront*), a second in the back of the robot (topic */scanback*) and a third one tilted 30° in the front of the robot (topic */scanvtcal*) for negative height obstacles avoidance and better anticipation of positive obstacles. These topics follow the datatype *sensor_msgs/LaserScan*[5]. Both frontal and back lasers publish data at a rate of 40 *Hz*, while tilted one does at 10 *Hz*.
- Transformations between sensors. Topic */tf* of datatype *tf/tfMessage*[6] offers information about static transformations between sensors and robot system reference, as sensors are fixed to robotic platform, and also transformation between starting point (frame_id: /odom) and robot (frame_id: /base_link). Frame_id's of sensors are: /bumblebee (camera), /laserfront (frontal laser), /laserback (back laser) and /laservtcal (tilted laser). This information is published at a rate of 10 *Hz* each.
- Camera Info. Topics */bumblenode/left/camera_info* and */bumblenode/right/camera_info* of datatype *sensor_msgs/CameraInfo*[5] carry information about camera calibration. They contain the camera calibration for left and right cameras using ROS interface. They are published synchronized with the images at 5 *Hz* approximately.
- Raw RGB image. Topics */bumblenode/left/image_raw* and */bumblenode/right/image_raw* of datatype *sensor_msgs/Image*[5] with the images gathered form the stereo pair. The images are captured at 5 *Hz* approximately.
- Rectified image. Topics */bumblenode/left/image_rect* and /bumblenode/right/image_rect of datatype *sensor_msgs/Image*[5], this time grayscale rectified images. They have been computed offline, the rectified images and camera info contain the same time stamp as the source raw image.

III. HUMAN - ROBOT INTERACTION EXAMPLES

The datasets presented in this work represent a valuable set of images and range measurements that could be used to the comprehension of several human-robot behaviours or interactions. Lengthwise for more than $4h51m30s$ of collected video, a great collection of real interaction scenes could bring us a qualitative idea of the different behaviours by studying and tagging it.

For the evaluation of the HRI experiment, it is important to notice that the experiments were made under the premise of show the same human-pilot-behavior in all trials. In addition, the robot appearance has to be into account (uncanny valley). Next lines summarize the HRI setup:

- People were not explicitly advertised about the presence of the robot.
- Same human-pilot at all experiments.
- The pilot always try to stay far away from the robot platform in order to not influence into people's reaction.
- The goal was to perform every day a navigation across several pre-established places at the Alcazar.
- The maximum linear speed was 0.5*m/s*.
- The pilot always try to perform every single experiment in the most fast and polite way.
- If any person try to stop the robot for some reason, the pilot tried to avoid that situations without personally taking part, only with the help of robot platform movements to demonstrate the intention of scape.

Research in human-robot interaction shows that autonomous social-capable robot navigation must considers the following behaviors:

- Respect personal zones
- Respect affordance spaces
- Avoid culturally scorned upon behaviors
- Avoid erratic motions or noises that cause distraction
- Reduce velocity when approaching a person
- Approach from the front for explicit interaction
- Modulate gaze direction

This list only represents the research so far, but it is possible that in the future new human-aware capabilities are added to that list. The datasets that we present here represent a good example for them, although placing a robot in a human environment also requires a special form of movements, like follow a person, solve blocked passage and dense crowd, guide a person/group of persons (not present at the experiments) or move in formations.

[5]http://wiki.ros.org/sensor_msgs
[6]http://wiki.ros.org/tf

TABLE I

PLOT AND DESCRIPTION OF THE DIFFERENT DATASETS. BLACK ASTERISK: STARTING POINT, RED ASTERISK: ENDING POINT. RED ASTERISK MAY OVERLAPS BLACK ONE, DUE TO PROXIMITY BETWEEN ENDING AND STARTING POINTS.

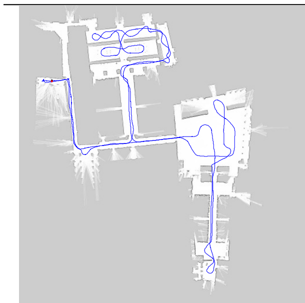
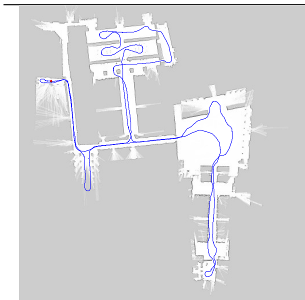
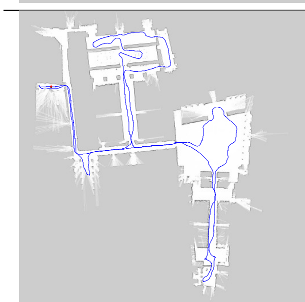
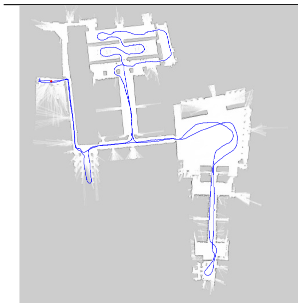
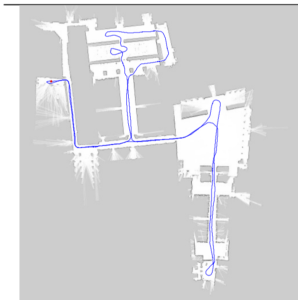
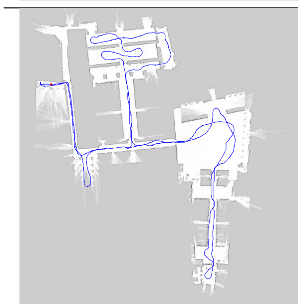
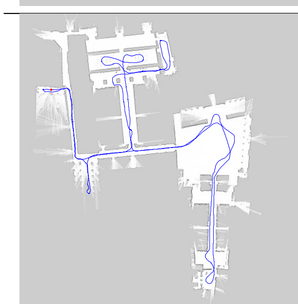
Robot Path	Dataset name	Description
	Dataset 1 8:41am	The robot traveled *1899.88 m* for *30:15s* receiving 8170 images and 72644 range measurements (frontal laser).
	Dataset 2 9:44am	The robot traveled *1820.02 m* for *29:10s* receiving 7876 images and 70046 range measurements (frontal laser).
	Dataset 3 10:31am	The robot traveled *1666.00 m* for *26:42s* receiving 7212 images and 64238 range measurements (frontal laser).
	Dataset 4 11:36am	The robot traveled *1845.66 m* for *29:41s* receiving 8015 images and 71417 range measurements (frontal laser).
	Dataset 5 12:43pm	The robot traveled *1970.69 m* for *31:43s* receiving 8567 images and 76088 range measurements (frontal laser).
	Dataset 6 14:57pm	The robot traveled *1824.06 m* for *29:09s* receiving 7870 images and 70062 range measurements (frontal laser).
	Dataset 7 15:53pm	The robot traveled *1569.47 m* for *2516s* receiving 6825 images and 60758 range measurements (frontal laser).
	Dataset 8 16:41pm	The robot traveled *1843.57 m* for *29:29s m* receiving 7963 images and 70923 range measurements (frontal laser). Back laser crashed after 914s of execution.
	Dataset 9 17:40pm	The robot traveled *1894 m* for *30:16s* receiving 8176 images and 72833 range measurements (frontal laser).
	Dataset 10 18:37pm	The robot traveled *1864.08 m* for *29:57s* receiving 8087 images and 71945 range measurements (frontal laser).

Fig. 3. (a) Camera occlusion; Blocked gates: (b) still standing people, (c) walking people; Crossing people: (d) no socially conventional way (by the left), (e) socially conventional way; (f) Approaching people; (g) Crowded corridors ; (h) Crossing people; (i) Following groups of people ; (j) Close interaction

Qualitative information could be extracted from the visualization of the videos or the images to propose new algorithms or techniques to tackle this kind of problems. Furthermore, we could have a rich set of examples even in changing situations (avoid people in more narrow or wide corridor, or maybe at big squares; it is also possible to find examples of avoiding people at a corridor which perimeter is a vegetable fence,...). For some examples of detailed scenes, please refer to Fig. 3. On the other hand, all the data is available as images and range measurements, both from onboard sensors (like we described before). This kind of data is useful for many applications, such as validating techniques about people detection and tracking (from camera visual [11] or range laser measurements, as well as if we are interested for individual [2] or groups [13]).

Finally, this rich information source could be used to classify and to identify the observe behaviors. We might extract some of the human-robot interaction features from the sensors mounted into the mobile robot. It turns out that, in contrast to [3], [20], [18], [22], [7], the training data needs to be captured while the robot is actually present.

IV. LOCALIZATION

From the point of view of localization, the scenario considered in this work is a crowded mostly planar area. Some ramps are present in robot's trajectory that can be eventually detected as obstacles by horizontal lasers, affecting to localization and navigation. Tilted laser is added to be used for small steps and ramps detection and determine by software if the robot will be able to cross the area or not.

Most traveled area presents good scan matches even when the robot is partially surrounded by people (see Table I). But, as this scenario is highly crowded, the robot may become complete surrounded and lost or converged to a wrong location. To solve this situation, images gathered with both left and right cameras make the dataset interesting for testing loop closure and kidnapping algorithms based in scene recognition.

Having a nearly full day dataset allows training algorithms for different situations or illumination variances, as shown in Fig. 4. These datasets offer the possibility to study daylight variation in a mixed indoor outdoor structured environment and use this information for improving localization or augmenting accuracy.

This dataset is also useful for testing algorithms for removing illumination effect in scenes [5], [14], [16] destined to provide outdoor long-term localization based on scene recognition for mobile robots, because of the recording of illumination variation in same scenes over the range of time covered by datasets.

V. CONCLUSIONS AND FUTURE WORK

In this paper we present a set of datasets collected at Royal Alcázar of Seville at 10 different hours of the day. As explained above, this is a crowded mostly planar scenario with mixed indoor and outdoor areas.

These datasets present a valuable HRI information according to the used setup. As we mentioned before, they provide a large set of scenes which could be labeled or tagged, and after that the data could be analyzed by range measurements and/or image processing to propose models of human behaviors, predictions, etc. It is also a great source of qualitative information that could help researchers to prepare HRI experiments.

Fig. 4. Strong daylight variations at Royal Alcázar of Seville, that may cause in the same scene different features detection in visual algorithms.

The data is very interesting for localization purposes, particularly for testing long-term localization algorithms based in rangefinders or image processing. Nearly 20 *km* of data was gathered, with information of lasers, camera and odometry stored in ROS format. Commented repeatability is an added value to raw data, through the possibility to analyze algorithm responses in many tests with people and illumination variations in similar trajectories.

Future work will consider increasing the dataset with the position of persons in lasers and images, simplifying many applications such as human behavior learning and adaptation or to act as ground-truth for person detection algorithms.

References

[1] Kai O. Arras and Oscar Martinez Mozos, editors. *Special Issue on: People Detection and Tracking*, volume 2. International Journal of Social Robotics, March 2010.

[2] Kai Oliver Arras, Boris Lau, Slawomir Grzonka, Matthias Luber, Óscar Martínez Mozos, Daniel Meyer-Delius, and Wolfram Burgard. Range-based people detection and tracking for socially enabled service robots. In Erwin Prassler, Johann Marius Zöllner, Rainer Bischoff, Wolfram Burgard, Robert Haschke, Martin Hägele, Gisbert Lawitzky, Bernhard Nebel, Paul-Gerhard Plöger, and Ulrich Reiser, editors, *Towards Service Robots for Everyday Environments*, volume 76 of *Springer Tracts in Advanced Robotics*, pages 235–280. Springer, 2012.

[3] Maren Bennewitz, Wolfram Burgard, Grzegorz Cielniak, and Sebastian Thrun. Learning motion patterns of people for compliant robot motion. *I. J. Robotic Res.*, 24(1):31–48, 2005.

[4] José-Luis Blanco, Francisco-Angel Moreno, and Javier González. A collection of outdoor robotic datasets with centimeter-accuracy ground truth. *Autonomous Robots*, 27(4):327–351, November 2009.

[5] Peter Corke, Rohan Paul, Winston Churchill, and Paul Newman. Dealing with shadows: Capturing intrinsic scene appearance for image-based outdoor localisation. In *Proc. of the International Conference on Intelligent Robots and Systems (IROS)*, November 2013.

[6] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In Cordelia Schmid, Stefano Soatto, and Carlo Tomasi, editors, *International Conference on Computer Vision & Pattern Recognition*, volume 2, pages 886–893, INRIA Rhône-Alpes, ZIRST-655, av. de l'Europe, Montbonnot-38334, June 2005.

[7] G. Ferrer, A. Garrell, and A. Sanfeliu. Robot companion: A social-force based approach with human awareness-navigation in crowded environments. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pages 1688–1694, Nov 2013.

[8] Dorian Galvez-Lopez and J. D. Tardos. Bags of binary words for fast place recognition in image sequences. *IEEE Transactions on Robotics*, 28(5):1188–1197, October 2012.

[9] Arren Glover, Will Maddern, Michael Milford, and Gordon Wyeth. FAB-MAP + RatSLAM: Appearance-based SLAM for Multiple Times of Day. In *ICRA*, Anchorage, USA, 2010.

[10] Arren J. Glover, William P. Maddern, Michael Warren, Stephanie Reid, Michael Milford, and Gordon Wyeth. Openfabmap: An open source toolbox for appearance-based loop closure detection. In *ICRA*, pages 4730–4735. IEEE, 2012.

[11] Martijn Liem and Dariu M. Gavrila. Multi-person localization and track assignment in overlapping camera views. In Rudolf Mester and Michael Felsberg, editors, *DAGM-Symposium*, volume 6835 of *Lecture Notes in Computer Science*, pages 173–183. Springer, 2011.

[12] M. Luber, L. Spinello, J. Silva, and K.O. Arras. Socially-aware robot navigation: A learning approach. In *IROS*, pages 797–803. IEEE, 2012.

[13] Matthias Luber and Kai Oliver Arras. Multi-hypothesis social grouping and tracking for mobile robots. In Paul Newman, Dieter Fox, and David Hsu, editors, *Robotics: Science and Systems*, 2013.

[14] Will Maddern, Alex Stewart, Colin McManus, Ben Upcroft, Winston Churchill, and Paul Newman. Illumination invariant imaging: Applications in robust vision-based localisation, mapping and classification for autonomous vehicles. In *Proceedings of the Visual Place Recognition in Changing Environments Workshop, IEEE International Conference on Robotics and Automation (ICRA)*, Hong Kong, China, May 2014.

[15] Barbara Majecka. Statistical models of pedestrian behaviour in the forum. *Master's thesis, School of Informatics, University of Edinburgh*, 2009.

[16] Colin McManus, Winston Churchill, Will Maddern, Alex Stewart, and Paul Newman. Shady dealings: Robust, long- term visual localisation using illumination invariance. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Hong Kong, China, May 2014.

[17] Stefano Pellegrini, Andreas Ess, Konrad Schindler, and Luc van Gool. You'll never walk alone: Modeling social behavior for multi-target tracking. In *International Conference on Computer Vision*, 2009.

[18] Satoru Satake, Takayuki Kanda, Dylan F. Glas, Michita Imai, Hiroshi Ishiguro, and Norihiro Hagita. How to approach humans?: Strategies for social robots to initiate interaction. In *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction*, HRI '09, pages 109–116, New York, NY, USA, 2009. ACM.

[19] M. Smith, I. Baldwin, W. Churchill, R. Paul, and P. Newman. The new college vision and laser data set. *The International Journal of Robotics Research*, 28(5):595–599, May 2009.

[20] Simon Thompson, Takehiro Horiuchi, and Satoshi Kagami. A probabilistic model of human motion and navigation intent for mobile robot path planning. In Gourab Sen Gupta and Subhas Chandra Mukhopadhyay, editors, *ICARA*, pages 663–668. IEEE, 2009.

[21] Atsushi Yamashita Yusuke Tamura, Yoshitaka Terada and Hajime Asama. Modelling behaviour patterns of pedestrians for mobile robot trajectory generation. In Lazaros Nalpantidis Pablo Gonzalez-De-Santos and Alejandra Barrera, editors, *International Journal of Advanced Robotic Systems*, volume 11 of *Latest Trends in Mobile Robotic Research*. OPEN ACCESS Journal, 2013.

[22] Brian D. Ziebart, Nathan Ratliff, Garratt Gallagher, Christoph Mertz, Kevin Peterson, J. Andrew (Drew) Bagnell, Martial Hebert, Anind Dey, and Siddhartha Srinivasa. Planning-based prediction for pedestrians. In *Proc. IROS 2009*, October 2009.