The Marulan Data Sets:
Multi-Sensor Perception in Natural Environment with Challenging Conditions

Thierry Peynot
ARC Centre of Excellence for Autonomous Systems
Australian Centre for Field Robotics (ACFR)
The University of Sydney, NSW 2006, Australia
{tpeynot, scheding}@acfr.usyd.edu.au

Steve Scheding
Sami Terho
Department of Automation and Systems Technology
Aalto University, P.O.Box 15500
00076 Aalto, Finland
sami.terho@tkk.fi

Abstract
This paper presents large, accurately calibrated and time-synchronised data sets, gathered outdoors in controlled and variable environmental conditions, using an unmanned ground vehicle (UGV), equipped with a wide variety of sensors. These include four 2D laser scanners, a radar scanner, a colour camera and an infrared camera. It provides a full description of the system used for data collection and the types of environments and conditions in which these data sets have been gathered, which include the presence of airborne dust, smoke and rain.

1 Introduction
This paper presents a singular effort that has been made to constitute multi-sensor data sets to evaluate and compare perception algorithms for unmanned ground vehicles (UGVs), in particular in challenging environmental conditions. This data gathering was realised in a rural environment at the University of Sydney’s test facility near Marulan, NSW, Australia. This article is organised as follows. The first section is a short introduction to discuss the motivation for this work and the experimental design that was adopted. Section 2 presents the platform used for the data gathering, describing in particular its sensors and the calibration processes involved. Section 3 presents the collected data sets, illustrating in particular the environment types and the corresponding environmental conditions. Finally, Section 4 draws some conclusions, including suggestions for exploitation of these data.

1.1 Motivation
Public data sets are extremely useful to evaluate the performances of algorithms and to compare the results obtained by related work based on the same reference data. Some notable examples are:

- the Radish repository (Howard and Roy, 2008), featuring numerous logs of odometry, laser and sonar data, as well as maps, acquired mainly in indoor environments;
- the Victoria Park data set (Nebot, 2000), previously collected and published by the Australian Centre for Field Robotics (ACFR), which has been extensively used since 2001 to evaluate the performance of SLAM algorithms;
- the MIT Darpa Urban Challenge public data (Leonard et al., 2008), containing the logs of the MIT vehicle, including camera images and Velodyne 3D Lidar point clouds, in an urban environment;
- the New College Vision and Laser Data Set (Smith et al., 2009), composed of colour images and laser data collected in an urban environment (a campus and a park);
- the Rawseeds database (Matteucci et al., 2009), proposing data sets acquired mainly with cameras and laser range finders, outdoors in a campus and indoors, with ground truth.

However, there are too few examples of publicly available data sets gathered with a wide variety of sensors, in particular in outdoor natural environments, due to the time and financial cost involved in their acquisition. Bringing such data sets to the public not only provides a common reference to numerous researchers to evaluate their algorithms, it also provides real experimental data to research teams which may not have the necessary equipment at hand.

The main differences between the Marulan data sets presented in this paper and the aforementioned data sets are twofold. First, although sensors like stereovision bench are not included, more sensing modalities are proposed than in most public data sets, in particular with the presence of a radar and an infrared camera in addition to the commonly used lasers and visual cameras. Second, the data were acquired outdoors in natural and semi-urban environments and numerous data sets specifically contain challenging environmental conditions, while most data sets only propose nominal conditions in which the performance of perception is reasonably well understood.

1.2 Experimental Design
The data sets described in this paper have been collected especially for the general purpose of testing various perception algorithms (e.g. obstacle avoidance or terrain interpretation) for UGVs, with no very specific algorithm in mind, to limit the experimenter’s bias (i.e. the bias towards results expected by the human experimenter, typically the algorithm developer, who has expert knowledge of the technique under evaluation). In particular, conditions that are known to be problematic for the perception of UGVs were not avoided. On the contrary, they were
specifically included in this work (see Section 3.3), since they represent some of the most significant challenges for future work on perception. A common example of such challenging conditions is the presence of airborne dust. Indeed, it typically causes many state-of-the-art perception systems to fail, as noted in the CMU PerceptOR program outcomes (Kelly et al., 2006) and the DARPA Urban Challenge (Urmson et al., 2008). In that respect, this work can also be seen as a first step into promoting integrity in perception systems (Peynot et al., 2009b).

2 System Description

The vehicle used to collect the data is the 8 wheel skid-steering Argo platform (see Fig. 2), that has been retrofitted with sensors and actuators to make it a UGV. Typical operating speeds of the Argo in the context of this work are \(1 - 2\, \text{m/s}\). Its angular velocity is controlled through the brake pressures on both sides. This section describes the sensors mounted on this vehicle, the calibration of these sensors, and the time-synchronisation of the collected data.

2.1 Sensors

The following exteroceptive sensors were mounted on a common frame of the vehicle (see Fig. 1):

- four 2D Sick laser range scanners, which specifications are given in Table 1. Referring to the names shown in Fig. 1, the configurations of these lasers were the following.
  - LaserHorizontal was horizontally centered on the sensor frame, slightly pointing down to the ground (a few degrees of pitch), with zero roll.
  - LaserVertical was horizontally centered on the sensor frame, with 90 degree roll (thus scanning vertically) and zero pitch.
  - LaserPort was located on the port side of the vehicle, slightly pointing down to the ground (a few degrees of pitch) and zero roll.
  - LaserStarboard was on the starboard side of the vehicle, with both pitch and roll angles close to zero;

- a 94 GHz Frequency Modulated Continuous Wave (FMCW) Radar (custom built at ACFR for environment imaging), whose specifications are given in Table 1.
- a mono-CCD colour camera equipped with a 6.5mm lens, acquiring images at a nominal framerate of 15 frames per second \((fps)\) in static\(^1\) data sets and 10\(fps\) in dynamic data sets\(^2\) (see Table 2);
- a thermal infrared (IR) camera, with a spectral response range of \(7 - 14\, \mu m\). The infrared images were obtained through a frame grabber (see details in Table 2).

The vehicle was also equipped with a number of proprioceptive sensors, providing information such as wheel angular velocities, engine rotation rate and brake pressures.

The navigation solution was provided by a commercial off-the-shelf Novatel RTK DGPS/INS SPAN\(^3\) unit, composed of a Novatel ProPak-G2plus GPS receiver and a Honeywell HG1700 AG17 Inertial Measurement Unit (IMU). This usually provides a 2cm accuracy localisation, with a constant update of the estimated uncertainties on this solution. This navigation solution was output at \(50\, \text{Hz}\), in double precision.

The frames used in this work are illustrated on Fig. 2. They are defined as follows:

- The Navigation frame is a fixed global frame defined by the three axes: \(X^n = \text{North}, Y^n = \text{East} \) and \(Z^n = \text{Down}\) in which positions are expressed in UTM coordinates (Universal Transverse Mercator).

- The Body frame is linked to the body of the vehicle. Its centre is located at the centre of the IMU, approximately at the centre of the vehicle. The axes are: \(X^b\) pointing towards the front of the vehicle, \(Y^b\) pointing to the starboard side of the vehicle, and \(Z^b\) pointing down.

- A Sensor frame is linked to a particular sensor. Its axes are defined in a similar way as the previous one (i.e. \(X^s\) forward, \(Y^s\) starboard, \(Z^s\) down), but it is centered on the considered sensor.

\(^1\)see Section 3 for definition of static and dynamic in that context
\(^2\)a slightly reduced frame rate was used for the dynamic data sets to avoid data loss due to factors such as the vibrations of the system which affect the efficiency of the hard disk writing.
\(^3\)Synchronised Position Attitude & Navigation
2.3 Sensor Calibration

The spatial transformations between sensors and reference frames have been estimated using thorough calibration methods. Consequently, the sensor data are ready to be used to build 3D representations of the world and to achieve accurate multi-sensor data fusion. Two categories of calibration have been made: 

- a Range Sensor Calibration, to estimate the transformations between the Sensor frame associated to each range sensor and the Body frame,

- a Camera Calibration, to estimate the intrinsic geometric parameters of each camera, and the extrinsic parameters of the transformations between cameras and lasers.

2.3.1 Range Sensor Calibration

The estimation of the transformations between the frame associated to each range sensor (laser scanner or radar) and the Body frame was made using a technique detailed in (Underwood et al., 2010). It allows for joint calibration of multiple range sensors, to minimise systematic errors in individual sensors as well as the systematic contradiction between sensors. This is needed to perform low-level data fusion between all these sensors. For that purpose, a data set was acquired in an open area with key geometric features such as a flat ground, a vertical wall and a few vertical poles (see (Peynot et al., 2009a)). The calibration routine estimates the transformations between the Sensor frames and the common Body frame that minimise a global cost function representing how accurately the laser point clouds match the known geometry of the key features.

The outputs of this calibration are the estimation of the 3 rotation angles ($Roll_X$, $Pitch_Y$ and $Yaw_Z$) around the frame axes and 3 translation offsets ($dx$, $dy$, $dz$) that fully describe the transformation from the Body frame to the Sensor frame. Table 3 shows the results obtained after combined calibration of all four range sensors, i.e. LaserHorizontal (or LaserH), LaserVertical (or LaserV), LaserPort (or LaserP), LaserStarboard (or LaserS) and the Radar. All angles are expressed here in degrees for convenience, and distances in metres.

2.3.2 Camera Calibration

For each type of camera (IR and Visual) a two-step calibration was executed. It is described as follows. The first step is to estimate the intrinsic (geometric) parameters of each camera using the Camera Calibration Toolbox for Matlab (Bouguet, 2008). This requires a set of images featuring a chess board with squares of known dimensions. The second step then estimates the extrinsic transformations between cameras and lasers, using a method adapted from (Pless and Zhang, 2003). This method exploits the relation between the planar chess board surface as seen by the camera and the laser scanline on this same planar pattern. The same process was used for both colour and IR cameras. The only difference concerned the chess board. Indeed, for the calibration of the IR camera, a chess board had been printed on thick paper and stuck on a planar isolating material (8mm thick corrugated PVC sheet) using adhesive tape on the borders. It was then heated by exposing it to direct sunlight during the acquisition of the calibration images. The sizes of the black and white squares of these chess boards were the following:

- for the IR camera: 114.8mm on both sides,
- for the Visual camera: 74.9mm on the left-right axis as it can be seen in the images and 74.7mm on the axis corresponding to the direction up-down.

Note that the data sets (laser scans and images) that were used for these calibrations are provided, next to the multi-sensor data sets, so that the user can perform any other calibration method that relies on similar input data (same type of calibration features, in particular). The images in these data sets feature the chess board exposed with various orientations in space, and at various distances, as appropriate for the Matlab camera calibration toolbox that was used.

The results of the intrinsic calibration of both cameras can be found in (Peynot et al., 2009a). The estimated extrinsic parameters are given in this section. The offset translations ($\delta X_c$, $\delta Y_c$, $\delta Z_c$) and rotations ($\phi X_c$, $\phi Y_c$, $\phi Z_c$), indicated in Table 4 (Visual camera) and Table 5 (IR camera), describe how to move each laser so that it aligns with the camera. They are expressed in the camera frame, using the Matlab Toolbox convention (i.e. $+X_c$ to the right, $+Y_c$ down, $+Z_c$ forward, Fig. 3). In these tables...
Table 3: Transformations Body Frame to Sensor Frame

<table>
<thead>
<tr>
<th>Sensor</th>
<th>RollX</th>
<th>PitchY</th>
<th>YawZ</th>
<th>dX</th>
<th>dY</th>
<th>dZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>LaserH</td>
<td>-0.7328</td>
<td>-8.5869</td>
<td>-1.6313</td>
<td>0.1090</td>
<td>0.0083</td>
<td>-0.9197</td>
</tr>
<tr>
<td>LaserV</td>
<td>88.5630</td>
<td>-0.1180</td>
<td>-1.1231</td>
<td>-0.0003</td>
<td>-0.0823</td>
<td>-1.1268</td>
</tr>
<tr>
<td>LaserP</td>
<td>-0.5002</td>
<td>-2.6162</td>
<td>-1.8059</td>
<td>0.1909</td>
<td>-0.5488</td>
<td>-0.7638</td>
</tr>
<tr>
<td>LaserS</td>
<td>-0.6082</td>
<td>-0.4311</td>
<td>-2.3500</td>
<td>0.1987</td>
<td>0.5343</td>
<td>-0.8495</td>
</tr>
<tr>
<td>Radar</td>
<td>-0.1516</td>
<td>191.1617</td>
<td>173.2781</td>
<td>-0.0258</td>
<td>-0.0472</td>
<td>-1.3991</td>
</tr>
</tbody>
</table>

Table 4: Lasers to Visual Camera Transformations

- **LaserHorizontal to visual camera:**
  - \( \delta X_c \): 0.4139, \( \delta Y_c \): -0.2976, \( \delta Z_c \): -0.0099, \( \phi X_c \): -4.7341, \( \phi Y_c \): -0.3780, \( \phi Z_c \): -0.4230
- **LaserVertical to visual camera:**
  - \( \delta X_c \): 0.5045, \( \delta Y_c \): -0.0905, \( \delta Z_c \): -0.208, \( \phi X_c \): -13.2030, \( \phi Y_c \): -0.5851, \( \phi Z_c \): -88.3628
- **LaserPort to visual camera:**
  - \( \delta X_c \): 0.9592, \( \delta Y_c \): -0.5011, \( \delta Z_c \): -0.208, \( \phi X_c \): -10.6026, \( \phi Y_c \): -0.0747, \( \phi Z_c \): -0.5791
- **LaserStarboard to visual camera:**
  - \( \delta X_c \): -0.1343, \( \delta Y_c \): -0.4976, \( \delta Z_c \): -0.0532, \( \phi X_c \): -12.6652, \( \phi Y_c \): 0.2409, \( \phi Z_c \): -0.5293

Table 5: Lasers to IR Camera Transformations

- **LaserHorizontal to IR camera:**
  - \( \delta X_c \): -0.3391, \( \delta Y_c \): -0.3278, \( \delta Z_c \): 0.0975, \( \phi X_c \): -5.307, \( \phi Y_c \): -1.2671, \( \phi Z_c \): -2.1308
- **LaserVertical to IR camera:**
  - \( \delta X_c \): -0.2485, \( \delta Y_c \): -0.1207, \( \delta Z_c \): -0.0115, \( \phi X_c \): -14.9996, \( \phi Y_c \): -1.4742, \( \phi Z_c \): -90.0505
- **LaserPort to IR camera:**
  - \( \delta X_c \): 0.2090, \( \delta Y_c \): -0.5400, \( \delta Z_c \): 0.0194, \( \phi X_c \): -12.7686, \( \phi Y_c \): -1.0343, \( \phi Z_c \): -2.3348
- **LaserStarboard to IR camera:**
  - \( \delta X_c \): -0.8772, \( \delta Y_c \): -0.5652, \( \delta Z_c \): 0.0584, \( \phi X_c \): -15.7117, \( \phi Y_c \): -0.8259, \( \phi Z_c \): -3.3619

2.4 Time Synchronisation

As the UGV platform used in this work was equipped with several computers, the sensor data were coming from various sources with their own clocks. For example, all proprioceptive and navigation data were collected by the low-level control computer, running the real-time operating system QNX, while all laser data were gathered by the “Sensor Server” computer running Linux and the radar data were logged on a separate computer running QNX as well. To account for timing differences, all data were systematically time-stamped at the time of their acquisition, and NTP (Network Time Protocol) was used to reduce the time synchronisation errors between the computers to less than 3 ms.

3 The Data Sets

This section focuses on the actual data collected, describing in particular the two main types of data sets (static and dynamic), the natures of the perceived areas, and the controlled environmental conditions.

3.1 Static Tests

The static tests consisted of sensing a fixed ‘reference’ terrain, containing simple known objects, from a motionless vehicle (Fig. 4). The positions and sizes of all objects within the predefined test area being known (their measurements are given in Peynot et al., 2009a)), the actual geometry of the environment can be used as a ground truth (within the hand measurement error) to evaluate the ability of each sensor, or sensor combination, to accurately represent the environment.

3.2 Dynamic Tests

For these tests, data were acquired from a moving vehicle in three different areas, representative of typical UGV operating environment except for the radar data in sets 25 to 28, i.e. dynamic data sets at nighttime, where the synchronisation error with all other sensor data was less than 10 ms.
ments. The first area, named open area, was mainly composed of a large, roughly flat ground, delimited by fences, a metallic shed, a static car and a trailer. The Houses area comprised a few bar-
racks, a static car and a few trees and was delimited by fences. The Dam area featured trees, a static car, a couple of trailers and a small lake. Illustrations of these areas can be found with the
data. The average linear speed of the vehicle was 0.5m/s, with a maximum of 1.5m/s. Variable yaw rates were achieved, with a maximum of 1.12rad/s (i.e. 64°/s).

3.3 Controlled Environmental Conditions

For both categories, data have been gathered in controlled envi-
ronmental conditions, which included the presence of airborne dust, smoke and rain. Dust clouds were generated by blowing air across dry soil using a high-power air compressor. Smoke clouds were generated using emergency smoke bombs that worked for about one minute, and having the wind naturally carry them across the test area. Rain was generated using two different pro-
cedures. In the static tests, water was quite homogeneously and
continuously spread in the test area using sprinklers attached to the
top of the large metal frame seen in Fig. 4. However, in the dy-
namic test, rain was simulated by spraying water with a hand-held hose in front of the vehicle throughout the corresponding data set.

3.4 Summary

Let a particular data set consist of the continuous acquisition of
synchronised data from all sensors for a few minutes. The data
presented in this paper consist of a total of 40 separate data sets, in addition to 3 calibration-dedicated data sets, for a total amount of about 400GB of raw data. They are published with a technical report (Peynot et al., 2009a), describing all details on sensor characteristics, formats and content of files, at the following address:

http://sdi.acfr.usyd.edu.au/

4 Conclusion

In this work, large, accurately calibrated and synchronised, multi-
modal data sets, have been gathered in controlled environmental
conditions (including the presence of dust, smoke and rain) by a representative UGV equipped with various types of sensors. These data sets have been made available to the public to test and compare perception algorithms. This is all the more pos-
sible thanks to tests in a static environment where the sensors perceived a ‘reference’ scene with known objects geometry char-
acteristics that may be used as ‘ground truth’. Besides, while illustrating interesting and challenging cases for perception of on-
board UGVs, these data were gathered with no very specific al-
gorithms in mind, unlike most available data sets in the literature.
This significantly reduced the experimenter’s bias.
Possibilities of future work on perception exploiting these data sets are numerous. They include the promotion of sensor data integrity and reliable perception in outdoor environments, for 2D/3D terrain representation, obstacle detection or even pedes-
trian detection and tracking.

Acknowledgments

This work was supported by the US Air Force Research Labo-
atory through the Asian Office of Aerospace Research and De-
velopment (grant AOARD-08-4059); and the ARC Centre of Ex-
cellence programme, funded by the Australian Research Council (ARC) and the NSW State Government. The authors would also like to thank Craig Rodgers, Marc Calleja, James Underwood, Andrew Hill and Tom Allen for their valuable contribution to the gathering and processing of these data.

References


