The Cheddar Gorge Data Set

R. Simpson, J. Cullip and J. Revell
BAE Systems (Operations) Limited, Advanced Technology Centre
Sowerby Building, FPC 267, PO Box 5, Filton, Bristol, BS34 7QW, U.K.

January 24, 2011

Abstract

This paper presents the unique Cheddar Gorge data set intended for the multiple disciplinary research areas of ground based vehicle autonomy and field robotics. The name of the data set refers to a highly aesthetic and genuine canyon in the United Kingdom, Somerset, known as Cheddar Gorge, making a challenging yet exciting location to collect, share and use data from. Furthermore, the Cheddar Gorge data set makes several notable contributions, identified through extensive systems engineering and research. The most important contribution is that the criteria for selecting the sensors, platform and location were influenced directly from genuine customer requirements, from both civil and military application domains[1]. This has ensured that the data set is as uncontrived and unbiased as possible from any algorithm requirements, but instead favour real application conditions. In short, the Cheddar Gorge data set exploits a broad sensor payload consisting of a combination of active and passive sensors, deployed on a ground based vehicle in a challenging environment and made publicly available.

1 Introduction

The demand from research scientists to system engineers in the robotics community, for publicly available and peer-reviewed data sets of a high calibre peaked with the editorial, *Data Papers - Peer Reviewed Publication of High Quality Data Sets*, in the International Journal of Robotics Research [4]. The article was the first published in the Robotics field to recognise this demand and urged scientists to act. This demand was fundamentally driven by the need to support algorithm validation, performance benchmarking and system analysis.

Many of the most well cited data sets such as the Victoria Park [3], RADISH [2] and RAWSEEDS [1] have been publicised via the website OpenSLAM [9]. Additionally, there are the ACFR data sets [5] and as a result of the editorial [4], the New College [8] data sets, also both now publicly available. Typically, these provide small to medium scale data sets with multiple sensors and accurate ground-truth data, using differential global positioning systems (D-GPS) outdoors or a sensor fusion network indoors. However, all these have been influenced by each research group's own project requirements and available hardware resources.

In contrast, the need for the Cheddar Gorge data set emerged through extensive systems engineering in the field of autonomy. This system engineering approach comprised of capturing customer, user and system requirements and later systems analysis defining key performance indicators, metrics, bounds and tests. The process highlighted significant but often understated requirements that the Cheddar Gorge data set should meet. These included that the Cheddar Gorge data set:

1. Shall support a diverse range of state of the art algorithm research, from simultaneous localisation and mapping to scene classification.
2. Shall capture sensor data from a real environment representative of larger scale challenges.
3. Shall be independent from published and current research based algorithm requirements.

In the next section we introduce the platform, sensor suite and sensor configuration, with Sections 3 to 5 describing the sensor pose measurement, synchronisation and calibration processes. Then in Section 6 we describe the special location highlighting particular epochs of interest that occurred in the data set. This is followed by Section 7 which provides a set of figures for visualising samples of the different sensor data. The next three sections explain the known issues and most importantly how to gain access to the data set, concluding in Section 10 with our future aspirations.

---

[1] The customer in this case is BAE Systems (Operations) Limited, UK.

2 Platform, Sensor Suite and Sensor Configuration

The Cheddar Gorge data set used the BAE Systems, Advanced Technology Centre's Bowler Wildcat platform displayed in Figure 1. The Wildcat platform was chosen for being a large scale autonomous unmanned ground vehicle (UGV) demonstrator. Consequently, the Wildcat has genuine customer requirements for localising without the aid of GPS and provides numerous sensor mounting positions and computer hardware options. Where applicable, the sensors were mounted on the platform roof to meet legal and health and safety requirements.

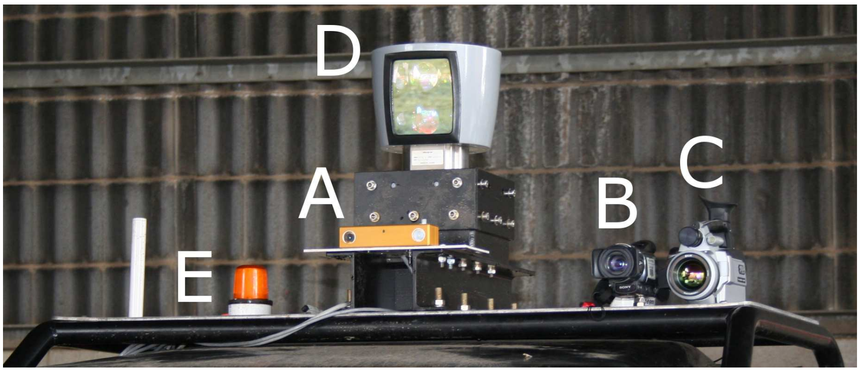


Figure 1: The BAE Systems, Advanced Technology Centre's Bowler Wildcat autonomous unmanned ground vehicle.

The selection and mounting of sensors was chosen to satisfy the first two requirements highlighted in Section 1. The sensor suite consisted of a:

- Point Grey Research stereo vision camera (model Bumblebee 2 BB2-08S2C-38) with two 1/3" progressive scan, charged couple devices (CCDs) mounted in a fixed canonical configuration, with a baseline of 12 cm. The mount position is shown in Figure 1 at label A.

- Sony high definition video camera (model HDR-HC1E) configured as a monocular forward facing camera, capturing interlaced 24 bit, MPEG2 compressed video with a resolution of 1920×1080 pixels. The mount position is shown in Figure 1 at label B.

- FLIR infra-red video camera (model ThermaCam P620) configured as a monocular forward facing camera, capturing 8 bit, MPEG2 compressed video with a resolution of 640×480 pixels. The mount position is shown in Figure 1 at label C.

- Velodyne 3D Light Detection and Ranging sensor (LiDAR) (model HDL-64E S2). This has a 360° horizontal field of view (FOV), a 26.8° vertical FOV and has a data rate of over 1.3 million point measurements every second up to a range of 120 meters. The mount position is shown in Figure 1 at label D.

- Oxford Technical Solutions inertial navigation system (INS) (model OXTS RT3044). This contains a US NAVSTAR GPS dual antenna receiver, local Differential GPS (DGPS) receiver, OmniSTAR subscription and high precision inertial measurement unit (IMU) providing acceleration and gyroscopic measurements. All localisation data is internally fused using a built in filter providing a ground-truth location estimate. The mount position is inside the vehicle above the rear differential.

- Xsens (model MTi-G-28) small and low cost IMU used to capture raw and uncalibrated, acceleration, magnetic field, gyroscopic and GPS data. The mount position is shown in Figure 1 at label E, behind the warning light.

- Set of four independent wheel distance encoders based on the anti-lock braking system (ABS). The raw voltage was captured from the ABS inductor coil located in each wheel using a Labjack U6-Pro analogue-to-digital converter (ADC) with 16-Bit precision at 25 kHz. Turning information can be calculated from the differences in wheel rotation speeds.

All sensor data was captured to a computer located in the rear of the vehicle. The computer hardware consisted of: a Dell PowerEdge R710 blade server, Dual Intel Xeon X5570 Quad Core processors at 2.93GHz, 16 GB of RAM and 1.5 TB of solid state disks in RAID 0 configuration.

3 Sensor Pose Measurements

In this section we provide the sensor pose measurements generated using a highly calibrated laser tracker device manufactured by Leica. This provides measurements to within a micron precision by averaging a total of 500 readings of a laser beam to a reflective target within line of sight. We determined a local coordinate system by measuring the position of a unique reference point on the Velodyne LiDAR located on the highest part of the vehicle. With this defined, we measured a central mounting point for each sensor relative to the reference point. A translation based on the technical drawings of each sensor finally yielded a relative 3D sensor position. The precision of the laser tracker allowed us to measure and calculate the 3D orientation of each sensor mounting plate. In total, the laser tracker can provide the full 6 degrees of freedom (DoF) for each sensor. To generate orientation, four points were measured on the plane surrounding the mounting position. Based on these measurements the laser tracker software calculated a roll, pitch and yaw relative to the reference point. In Figure 2 the coordinate systems for each sensor is shown.

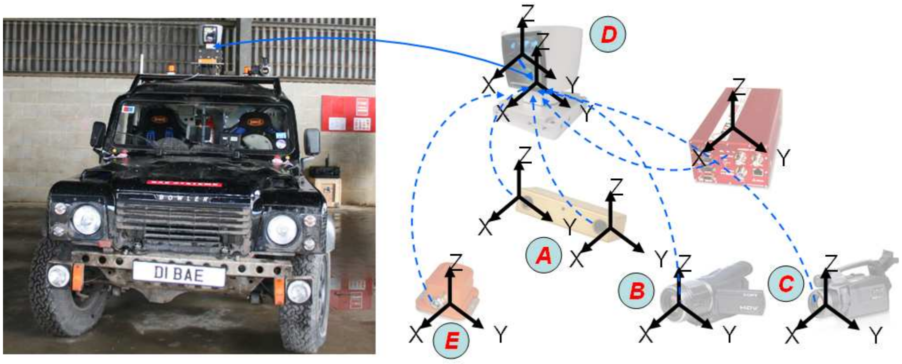


Figure 2: A schematic of the different coordinate systems used.

The 6 DoF sensor poses are provided in Table 1. These are based on the mean of 5 measurements to compensate for the human error associated with positioning the reflective target on the sensor mounting point.

Table 1: Sensor pose measurements.

Sensor Name	X (mm)	Y (mm)	Z (mm)	Roll X (rad)	Pitch Y (rad)	Yaw Z (rad)
Point Grey Bumblebee 2 (Left)	342	27	-139	0.0020	-0.0036	-0.0044
Point Grey Bumblebee 2 (Right)	342	147	-139	0.0020	-0.0036	-0.0044
Sony HDR-HC1E	394	499	-198	0.0236	0.0014	0.0013
FLIR ThermaCam P620	214	587	-208	0.0236	0.0014	0.0013
Velodyne LiDAR HDL-64E S2	89	89	-17	0.0005	-0.0005	0.0012
OXTS RT3044 INS	-948	174	-1018	-0.0262	1.4241	0.0426
Xsens MTi-G IMU	-111	-381	-252	-0.0086	-0.0082	0.0090
Wheel Rear Left	-1038	904	-1684	-	-	-
Wheel Rear Right	-1038	-666	-1684	-	-	-
Wheel Front Left	1662	904	-1684	-	-	-
Wheel Front Right	1662	-666	-1684	-	-	-

4 Sensor Calibration

We include the manufacturer's calibration data for the Xsens MTi-G IMU, Velodyne LiDAR and Point Grey Bumblebee 2. The calibration data for these sensors was not more than a month old. Wheel odometry is provided in the form of 4 voltage signals taken from the ABS inductor coils located in each wheel. The signal manifests as a sinusoidal wave with period and amplitude in proportion to the speed of each wheel. A full wave cycle indicates the movement of one ABS magnetic tooth moving past the inductor. Table 2 displays essential information to convert this signal into a vehicle pose estimate.

Table 2: Wheel odometry key parameters

Key	Value
Wheel Track	1580 ± 2 mm
Wheel Base	2700 ± 2 mm
Wheel Circumference	2520 ± 4 mm
Wheel Tread Width	180 ± 1 mm
ABS pulses per revolution	60

5 Sensor Synchronisation

Collecting data simultaneously from several sensors presents a significant challenge when attempting to synchronise data across multiple computers. In previously published data sets, sensor data was recorded to several computers, each synchronised using the Network Time Protocol (NTP) or IEEE 1588-2008 Precision Time Protocol (PTP) [1]. However, using network time synchronisation software will vary one or more clocks during the synchronisation process, causing a non-constant time difference between clocks. This results in imprecise time stamp data that can be detrimental to data fusion algorithms. Our approach to this problem was to avoid a multiple computer architecture. Instead, we relied on the latest computational hardware to simultaneously log sensor data onto a single device. This consisted of multiple core processors, an array of solid state disks in a RAID 0 configuration and a large RAM capacity, see Section 1. The available storage space and infrastructure of the Wildcat platform made the hardware installation and uninterrupted power requirements possible.

6 Location

Set in the Mendip Hills of Somerset, United Kingdom is the natural geological formation of Cheddar Gorge. Formed by glacial melt water over the last 1 million years, the gorge exhibits a steep cliff-face cut through the limestone hills. The gorge is surrounded by woodland and farmland, except to the southwest where the small village of Cheddar exists (at the lower end of the gorge). Along the centre of the gorge exists a carriage way allowing traffic to pass through providing a spectacular view of the 137m vertical cliff faces and rock formations (Figure 3). Cheddar Gorge was chosen for both its aesthetic topography and confined space which can provide interesting research challenges from both robotics and computer vision. Examples of such problems include 3D scene reconstruction; GPS multipath effects and moving object detection and tracking.

The vehicle platform and data set begin on the road to the east of the gorge. The vehicle traverses the gorge and completes a large loop via the villages of Cheddar and Draycott and a narrow country lane back to the start position. The gorge is passed a second time, after which a loop of Cheddar village is completed. Finally, the vehicle is driven back through the gorge in the reverse direction back to the start position. This course lasted just under 1 hour, capturing a total of 329 GB of sensor data over 31.8 km. Figure 4 shows the route taken by the vehicle as it drives through Cheddar Gorge, Cheddar Village and surrounding countryside.

7 Data Visualisation

In Figure 5, a visualisation of 4 out of the 7 sensors output is displayed. In this scene, a bus parked in the car-park gorge can be viewed simultaneously from (a) a visual camera, (b) a thermographic infra-red camera (c) a 3D LiDAR scan and (d) a stereo vision camera.

During the data set, there are several notable events that would be of interest to researchers. First, whilst moving south through the gorge, in a few instances a low sun angle causes extreme lens flare on the visual spectrum cameras. Second, there are many vehicles and pedestrians moving near the vehicle, both towards and with the direction of travel. Third, occasionally the GPS signal is lost on the OXTS INS causing an increase in the tracked position variance. This occurs when the vehicle passes through the gorge, or underneath tree foliage and provides an interesting case of GPS intermittence.

8 Data Issues

Capturing data sets of this scale, with this platform and with so many sensors is expensive. Therefore not every aspect of the Cheddar Gorge data set is perfect. In this section we highlight the issues and limitations of the data.

(a) (b) (c) (d) (e) (f)

Figure 3: Selected photographs of the Cheddar Gorge data set with the respective locations referenced in Figure 4

Unfortunately, the FLIR P620 infrared camera and Sony HD camera could not be integrated with the on-board computer. This meant that neither were synchronised to the other sensors. Further, the FLIR P620 paused recording intermittently to perform automatic radiometric adjustments. Consequently, as a result of neither sensor providing time stamp information, the data is difficult to accurately synchronise. Through experiments, we recommend that by using the known initial start time and by assuming a near constant frame rate of 25 fps, that the data is still extremely useful. Also , whilst the FLIR P620 infra-red camera provides extremely high quality thermal video imagery, see Figure 5.a, the images contain a small FLIR company logo and temperature bar embedded in every frame.

During the data set at some locations the OXTS RT3044 INS received signals from less than 4 GPS satellites. This caused occasional drifts in the localisation data. This happened when the vehicle was driven through the steep sided gorge and heavily wooded areas. This signal degradation was anticipated and is a characteristic of using these sensors for localising in these environments. It is also possible that the Xsens MTi-G IMU bias and gains recorded may be inaccurate based on limited testing immediately before data set collection. Also, the timestamps may be inaccurate from the IMU recording measurements in batch.

In addition to sensor clock synchronisation, discussed in Section 5, there is also the issue of the sensor delay time. This is the time difference between when the sensor captures the data and when the system clock time is recorded. Estimating this is non-trivial since there can be an unknown amount of processing on the local computer or on the sensor itself.

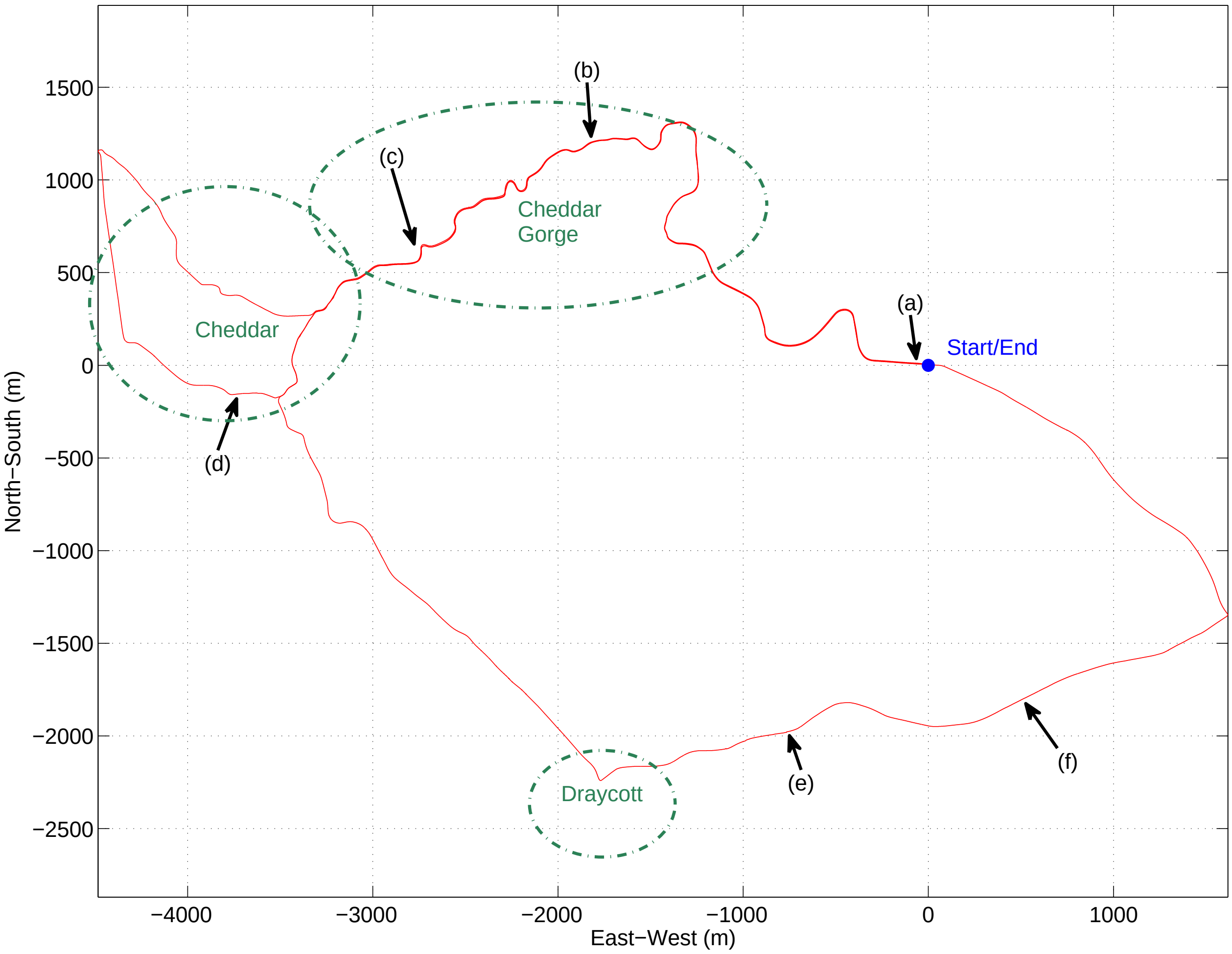


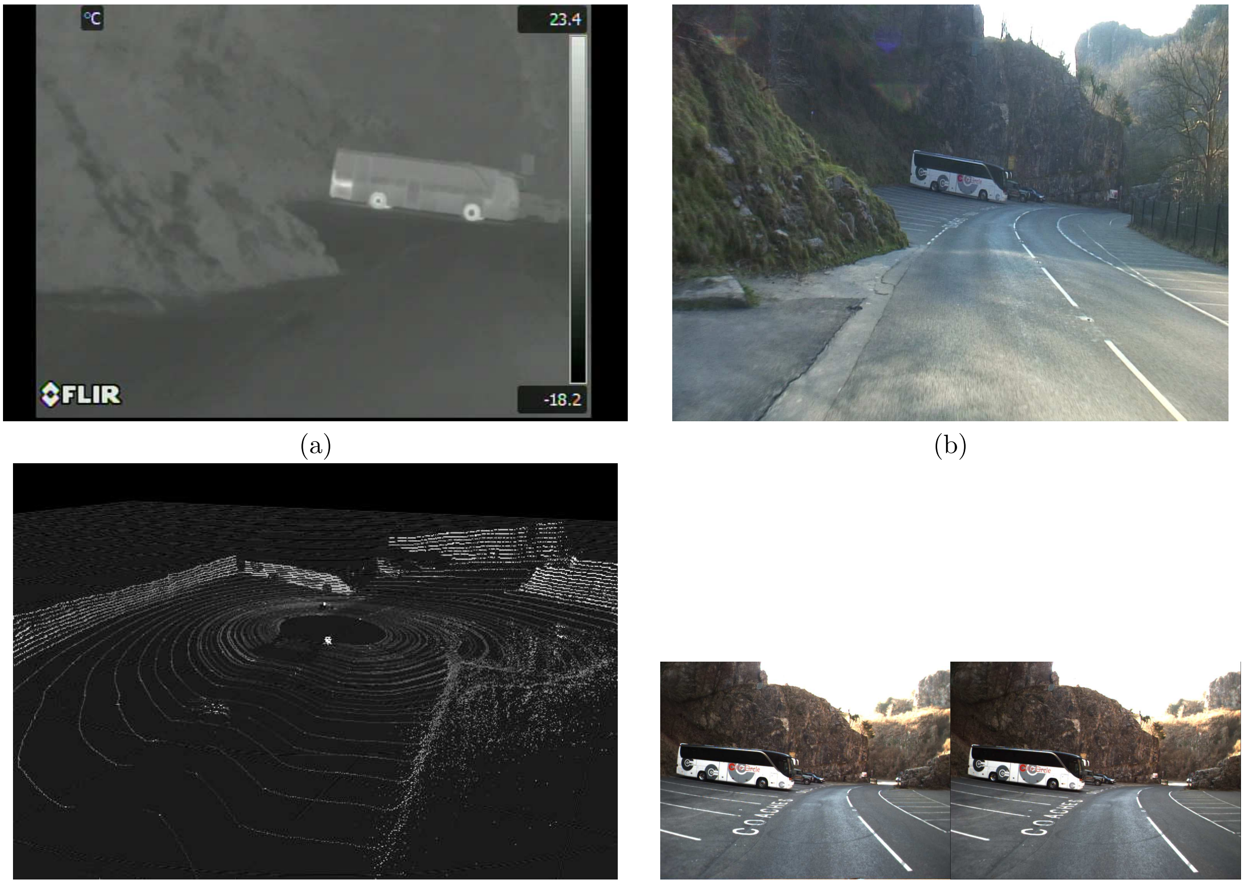
Figure 4: An aerial view of the vehicle trajectory with interest points referencing the photos in Figure 3

9 Distribution and Access

We are keen to make the Cheddar Gorge data set available to the research community under license. To apply for a copy, the reader must email *james.revell@baesystems.com* including the name of the researcher, organisation, address and brief statement on how they will use the data. Once received, a license and all additional information will then be posted to that address and a confirmation email sent. If the terms and conditions in the license are agreeable, the reader must send a signed copy of the license and an external hard disk drive (USB2, at least 350 GB to the address given on the license. When we have received the disk, the Cheddar Gorge data set will be copied onto it and sent back. Table 3 shows key information about the data set. Please note that no software will be provided to parse, visualise or process the data. Standard data formats are used throughout and data sheets are provided where applicable. References [7] and [6] should be used to cite this work.

Table 3: The Cheddar Gorge data set key information.

Data Set Name	Duration (minutes)	Average Speed (km/h)	Data Size (GB)	Environment Conditions	Lighting Conditions	Start Date & Time
Cheddar Gorge 2010	57	31.5	329	Dry, Sunny Clear & Cold	Low/Bright	5th March 2010 14:58 GMT



(a) (b)

(c) (d)

Figure 5: Sensor data captured in Cheddar Gorge from (a) FLIR P620 Infra-red camera, (b) Sony HD Camera and (c) Velodyne LiDAR point cloud and (d) the Point Grey Bumblebee 2 stereo camera.

10 Future

To adapt to future advances in technology, further data sets will be captured using a similar platform, with different sensor combinations, using updated solutions and meeting new system requirements.

Acknowledgements

This research was carried out by the Advanced Technology Centre (ATC) at BAE Systems (Operations) Limited and funded by the Systems Engineering for Autonomous Systems (SEAS) Defence Technology Centre (DTC) established by the UK Ministry of Defence (MOD), project reference AA022. The authors would also like to thank Chris Smith, Richard Grover, Glenn Callow, Gary Cross, Pete Day, Alex Stewart, Stephen Taylor and Kevin Hobbs at the ATC for providing support.

References

[1] A. Bonarini, W. Burgard, G. Fontana, M. Matteucci, D. Sorrenti, and J. Tardos. RAWSEEDS: Robotics advancement through web-publishing of sensorial and elaborated extensive data sets. In *IROS, Workshop on Benchmarks in Robotics Research*, 2006.

[2] A. Howard and N. Roy. The Robotics data set repository (RADISH), 2003.

[3] E. Nebot. The Sydney Victoria Park data set, January 2000.

[4] P. Newman and P. Corke. Editorial: Data papers - peer reviewed publication of high quality data sets. *International Journal for Robotics Research (IJRR)*, 28(5):587, May 2009.

[5] T. Peynot and S. Scheding. Datasets for the evaluation of multi-sensor perception in natural environments with challenging conditions. In *In Proceedings of Robotics: Science and Systems*, 2006.

[6] R. Simpson, J. Cullip, and J. Revell. The BAE Systems Wildcat Data Sets. In *5th Annual Technical Conference on Systems Engineering for Autonomous Systems (SEAS) Defence Technology Centre (DTC), July 2010, Edinburgh*, page C5. UK MoD, July 2010. `http://www.seasdtc.com`.

[7] R. Simpson, J. Cullip, and J. Revell. The Cheddar Gorge Data Set. Technical Report, Advanced Technology Centre, BAE Systems (Operations) Ltd, Sowerby Building, FPC 267, PO Box 5, Filton, Bristol, BS34 7QW, U.K., 2010. `http://www.openslam.org`.

[8] M. Smith, I. Baldwin, M. Churchill, R. Paul, and P. Newman. The New College vision and laser data set. *International Journal for Robotics Research (IJRR)*, 28(5):595–599, May 2009.

[9] C. Stachniss, U. Frese, and G. Grisetti. OpenSLAM, January 2009.