

2.3.1 Sufficiency of training

Practical guidance –automotive

Authors: Dr Matthew Gadd and Dr Daniele Martini, University of Oxford

Operating scenarios in academic and commercial research for Autonomous Vehicles (AVs) focus on structured, human-built environments. Although these scenarios form important and obvious use-cases, we must also include the vast scope of natural scenery crucial for industries such as construction, agriculture, mining, entertainment and forestry [1, 2]. Here, hazards span traditional, many-participant traffic environments to those featuring treacherous surfaces and rare-but-critical perceptual scenarios in rough outdoor deployments. Throughout, we emphasise the complementary characteristics of conventional and emerging sensing technologies in terms of suitability to these different scenarios [3].

Dataset overview

Our key contribution is the deployment of one platform in five places ranging from on-road to off-road environments. Our focus for data capture revolves around unusual sensing modalities, mixed driving surfaces, varied operational domains, and adverse weather conditions [4]. We, therefore, complete thousands of kilometres of driving in both rural and urban scenes across England and Scotland with our Range Rover RobotCar. Our sensor suite comprises sensors traditionally exploited for AVs, including cameras, LiDARS, and GPS/INS. However, we also include uncommonly-used sensors that show great promise in challenging scenarios, including FMCW radar and audio. Finally, we derive rare and unusual annotations (audio narration, etc.) alongside the typical vision-based ground-truth approaches.



Figure 1. Vehicle platform and sensor suite and the location in the UK for the collection sites.

The dataset is collected using our Range Rover RobotCar platform. This vehicle is ideal for mixed driving conditions - on and off-road - and has load and people carrying capacity. For these reasons, it is representative of and suitable for a large swathe of possible deployment scenarios.

The vehicle is equipped with an array of sensors, including:

- A forward-facing stereo camera, typically applied to motion estimation and localisation [5], and a set of three two-lens stereo cameras facing forwards and backwards, for around-vehicle object detection and semantic scene understanding [6]. Cameras are popularly applied but susceptible to appearance changes in scenarios of poor weather [7].
- A set of five single-beam, 2D lasers and a single multiple-beam 3D laser which form the other primary sensor technology applied to navigation [8] and scene understanding [9]. These sensors are more robust to illumination but also suspect to extreme weather scenarios [10].
- A roof-mounted scanning radar, increasingly applied to navigation [11, 12, 13] and scene understanding [14, 15, 16] as it is inherently immune to weather or illumination.
- A bumper-mounted automotive radar, more classically applied to object detection and already deployed extensively in the semi-autonomous market [17].
- An Inertial Navigation System (INS) for proprioceptive motion understanding in the case that perception systems fail and are commonly used alongside vision, laser, and radar sensors.
- An in-cabin monocular camera for recording driver behaviour both to advance driver assistance systems and for a rich source of information for supervising learnt methodologies.
- The internal controller area network (CAN) bus signals for directly learning how to drive in different scenarios, with a view towards linking this with perception systems such that decision-making for driving actions on a highly connected vehicle level is based on the perceived scenario.
- Four omnidirectional boom microphones on the two front and two optional back wheel arches to help terrain assessment [16, 18] in the case of total lack of illumination or light-based sensor degradation.
- An in-cabin, optional microphone to complement the in-cabin video under professional driving instruction and to enrich the situational awareness of our vehicles, particularly considering ground signatures and detection of other moving audio sources.

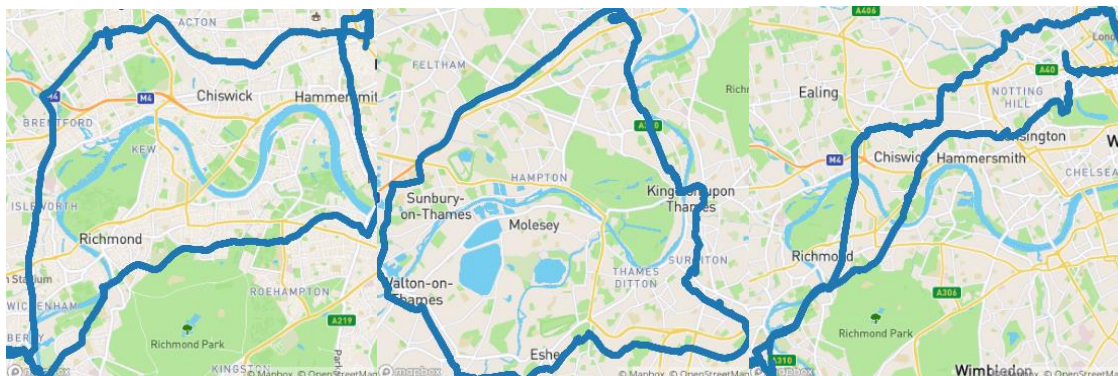


Figure 2. GPS traces overlaid on map data for the three different routes we traverse for data collection in Central and South-West London.

In collecting this dataset, the Range Rover RobotCar has visited various sites which we propose to be representative of important and difficult AV application scenarios. Aside from application potential, we view these as lying on a spectrum of scene types, in terms of (1) appearance, (2) presence of other actors and (3) driving difficulty.

- Hounslow Hall Estate, in Buckinghamshire, England. This site features three routes, mainly off-road, of increasing difficulty even under fair conditions. Additionally, any recent poor weather will make driving even more challenging (e.g. degrading the driving surface or the visibility of obstacles). The appearance is dense, visually ambiguous vegetation. This site is representative of operating scenarios requiring highly non-planar off-road driving, such as search-and-rescue.
- Ardverikie Estate, in the Scottish Highlands. This isolated estate features gravel and rock tracks and sandy beaches in highly variable lighting and weather conditions. This off-road scenario presents a distinct visual appearance and more diverse scenery (four routes) than the Hounslow Hall Estate (Scottish vs. English scenery, e.g. more regular fall/winter snow) but is slightly less challenging in terms of driving conditions. This site is representative of operating scenarios in natural heritage management.
- The New Forest, a large area of unenclosed pasture land and forest in Southern England. Along the spectrum of driving difficulty, this site presents easy surfaces (public roads) and low traffic. It has a visual appearance typical to the rural countryside, also featuring small villages. Despite the low traffic, there are other mobile vehicles in this area and - uniquely to this part of Hampshire and Wiltshire - productive animals (horses, cattle) roam the area freely. We see this site as representative of operating scenarios in rural activity.
- The Oxford Ring-Road, in Oxfordshire. This site features driving at high-speed (up to 70 mph) around a network of A-, B-, and M-type motorways. This driving is not difficult in the sense of surface (although some hazardous wet-weather conditions are captured), but it is an essential part of the dataset as we need AVs to perceive hazards even when travelling at high-speed (stopped cars, etc). Visually, this site presents semi-urban and motorway scenes. We see this data as being useful in application scenarios such as haulage and long-distance people transport.
- Central and South-West London, specifically around Shepherd's Bush, Notting Hill, Belgravia, East Sheen, etc. In this area, we perform dense urban driving with a professional driving instructor. Visually, this site gives us city scenery - including bicyclists, pedestrians, motorbikes, traffic control and signs, road markings, etc. Driving difficulty in this area can be complicated by the movement of many other actors and their shared use of roads. We see this site as an operating scenario suitable for daily short-distance people movement and congestion management.



Figure 3. Our platform pictured at the various dataset collection sites. From the top left: the Hounslow Hall Estate, the Ardverikie Estate, Central London and the New Forest.

Annotation

Aside from the terrain difficulty, the richness of appearance and structure for sensor recordings, and traffic conditions discussed above, each of our sites was also chosen based on the opportunity to harvest after data collection both popular as well as unusual and rare annotations for training machine learning systems. These include, on a site-by-site basis:

- Bounding box object positions, where the image region corresponding to an object are required but a dense scene decomposition is not (see semantic segmentation, below). These are specific to a class of object and will therefore dictate scenarios directly. For example, “the car in front is stopped despite the next nearest traffic light showing green”. These bounding boxes are annotated for camera as well as for LiDAR (where we provide 3D boxes) and radar (with a shorter list of categories, as radar blobs are less discriminative). For radar, in particular, there is a lack of such annotations available. Bounding box datasets abound in the literature [19, 20, 21], but we collect them for cameras in order to correspond them with audible explanations (see below). These annotations are harvested across primarily the Oxford Ring-Road and Central and South-West London, as these feature many more object types and actors than the off-road sites. Additionally, the New Forest site features both live traffic and freely roaming livestock.
- Pixel-wise segmentation, many datasets for which exist in urban driving [22, 23, 24]. However, for off-road driving, there is a paucity of road-surface labels (especially across multiple surface types) [25, 26]. This is important in deciding the style of driving which is most appropriate for a given scenario, and as such we also provide correspondences with CAN signalling (see above). These annotations are harvested primarily in the Ardverikie Estate and Hounslow Hall Estate sites.

- Event-based audio-visual sequences, where we uniquely provide synchronised audio and video. In urban sites (Central and South-West London) the synchronised audio is recorded in-cabin by a professional driving instructor who narrates all driving decisions. We provide correspondences with object positions (see above) which are mentioned in the narrative. In off-road sites (Ardverikie Estate and Hounslow Hall Estate) the synchronised audio is from microphones mounted in the wheel arches, for better appraisal of scenarios especially in poor light.
- Position tracking, using a Leica Viva TS16 Total Station for precise millimetre-accurate position ground truth for training systems to recover motion in highly non-planar scenarios.

We also make an effort to repeat some categories and label types across sites, for greater portability of learned knowledge when deploying to new scenarios. For example, driving surface segmentation and classification in Scotland also features tarmac.



Figure 4. Examples of annotations on the data. From the left: bounding boxes and pixel-wise segmentation in Central London, and pixel-wise segmentation of road surfaces in the Hounslow Hall Estate.

Summary

In summary, we present a universal view of AV operating scenarios. We would ideally like our vehicles to be deployable in any situation with limited time to develop capability. In order to achieve this we require (1) a complementary and robustly populated sensor suite, (2) datasets which feature driving scenarios as widely as possible, and (3) consistent annotation of all pertinent scenario elements and rich in the sense of linking sensors in the suite and being portable across scenarios.

References

- [1] Duckett, Tom, et al. "Agricultural robotics: the future of robotic agriculture." arXiv preprint arXiv:1806.06762 (2018).
- [2] Li, Jian-guo, and Kai Zhan. "Intelligent mining technology for an underground metal mine based on unmanned equipment." *Engineering* 4.3 (2018): 381-391.
- [3] Liu, Oi, Shihua Yuan, and Zirui Li. "A Survey on Sensor Technologies for Unmanned Ground Vehicles." 2020 3rd International Conference on Unmanned Systems (ICUS). IEEE, 2020.
- [4] Gadd, Matthew, et al. "Sense–Assess–eXplain (SAX): Building Trust in Autonomous Vehicles in Challenging Real-World Driving Scenarios." 2020 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2020.

- [5] Saputra, Muhamad Risqi U., Andrew Markham, and Niki Trigoni. "Visual SLAM and structure from motion in dynamic environments: A survey." *ACM Computing Surveys (CSUR)* 51.2 (2018): 1-36.
- [6] Hao, Shijie, Yuan Zhou, and Yanrong Guo. "A brief survey on semantic segmentation with deep learning." *Neurocomputing* 406 (2020): 302-321.
- [7] Garg, Kshitiz, and Shree K. Nayar. "Vision and rain." *International Journal of Computer Vision* 75.1 (2007): 3-27.
- [8] Jonnavithula, Nikhil, Yecheng Lyu, and Ziming Zhang. "LiDAR Odometry Methodologies for Autonomous Driving: A Survey." *arXiv preprint arXiv:2109.06120* (2021).
- [9] Gao, Biao, et al. "Are We Hungry for 3D LiDAR Data for Semantic Segmentation? A Survey of Datasets and Methods." *IEEE Transactions on Intelligent Transportation Systems* (2021).
- [10] Heinzler, Robin, et al. "Weather influence and classification with automotive lidar sensors." *2019 IEEE intelligent vehicles symposium (IV)*. IEEE, 2019.
- [11] Barnes, Dan, and Ingmar Posner. "Under the radar: Learning to predict robust keypoints for odometry estimation and metric localisation in radar." *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020.
- [12] De Martini, Daniele, Matthew Gadd, and Paul Newman. "kRadar++: Coarse-to-Fine FMCW Scanning Radar Localisation." *Sensors* 20.21 (2020): 6002.
- [13] Gadd, Matthew, Daniele De Martini, and Paul Newman. "Contrastive Learning for Unsupervised Radar Place Recognition." *arXiv preprint arXiv:2110.02744* (2021).
- [14] Kaul, Prannay, et al. "Rss-net: Weakly-supervised multi-class semantic segmentation with FMCW radar." *2020 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2020.
- [15] Broome, Michael, et al. "On the Road: Route Proposal from Radar Self-Supervised by Fuzzy LiDAR Traversability." *AI* 1.4 (2020): 558-585.
- [16] Williams, David, et al. "Keep off the grass: Permissible driving routes from radar with weak audio supervision." *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2020.
- [17] Bilik, Igal, et al. "The rise of radar for autonomous vehicles: Signal processing solutions and future research directions." *IEEE Signal Processing Magazine* 36.5 (2019): 20-31.
- [18] Zürn, Jannik, Wolfram Burgard, and Abhinav Valada. "Self-supervised visual terrain classification from unsupervised acoustic feature learning." *IEEE Transactions on Robotics* 37.2 (2020): 466-481.
- [19] Sheeny, Marcel, et al. "Radiate: A radar dataset for automotive perception." *arXiv preprint arXiv:2010.09076* 3.4 (2020): 7.
- [20] Behrendt, Karsten. "Boxy vehicle detection in large images." *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*. 2019.

- [21] Houston, John, et al. "One thousand and one hours: Self-driving motion prediction dataset." arXiv preprint arXiv:2006.14480 (2020).
- [22] Geyer, Jakob, et al. "A2D2: Audi autonomous driving dataset." arXiv preprint arXiv:2004.06320 (2020).
- [23] Liao, Yiyi, Jun Xie, and Andreas Geiger. "KITTI-360: A Novel Dataset and Benchmarks for Urban Scene Understanding in 2D and 3D." arXiv preprint arXiv:2109.13410 (2021).
- [24] Caesar, Holger, et al. "nusenes: A multimodal dataset for autonomous driving." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.
- [25] Metzger, Kai A., Peter Mortimer, and Hans-Joachim Wuensche. "A Fine-Grained Dataset and its Efficient Semantic Segmentation for Unstructured Driving Scenarios." 2020 25th International Conference on Pattern Recognition (ICPR). IEEE, 2021.
- [26] Wigness, Maggie, et al. "A rugd dataset for autonomous navigation and visual perception in unstructured outdoor environments." 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2019.