Pre-print of article that will appear in IEEE Potentials.

0

self-driving vehicles: key technical challenges and progress off the road

IEEE Potentials magazine theme issue on self-driving vehicles

Michael Milford, Sam Anthony and Walter Scheirer | January 5, 2020

1 Introduction

Self-driving vehicles, also referred to as autonomous vehicles (AVs) or driverless cars, have become in a period of less than 10 years one of the biggest technological arms races in the world, with tens of billions of dollars poured into companies and start ups [1]. The crown jewel in the race is the on road, consumer driverless car: whether owned by individuals or part of a centralized ride-sharing fleet, this is the area where the majority of investment has occurred. Yet autonomous vehicles have been around for much longer in other fields such as mining, domains which share some but not all of the same technical challenges faced by on-road autonomous vehicles.

In this article, we provide an overview of the key technical challenges and solutions for both on- and off-road autonomous vehicles, with a focus on one of the key unsolved challenges - interaction with vulnerable road users.

2 Key Technical Competencies - Hardware

A typical autonomous vehicle contains a number of key components: the physical platform itself, and a suite of sensing and onboard computational hardware.

2.1 Platform

The type of autonomous vehicle platform affects the viability of different technological solutions to autonomy. Larger vehicles are typically heavier and harder to stop, and more damaging when they hit something, but can carry more onboard sensing and compute. Energy storage also generally scales favourably with vehicle size, an important consideration which can enable better up-time percentages and utilization of more power hungry compute.

2.2 Sensor Suites

Figure 1: A typical sensor suite on top of a car, with multiple cameras and LIDAR sensors. Photo from QUT.

Autonomous vehicle platforms have access to a range of sensing technologies. LIDAR and laser-based range sensors provide accurate long distance "range to object" information, and can also use reflectance information to detect lane markings. In adverse weather such as rain or smoke, their capabilities can be significantly degraded.

Modern camera technology provides very high resolution imagery of the environment, with good dynamic range (revealing detail both in bright and dark areas of an image simultaneously) and high frame rates. The information present in a camera image is much richer than that provided by any other sensing modality, provided it can be successfully extracted: the widely quoted proof of concept here being that humans can drive very well with primarily visual sensing alone. Cameras are often cheaper and require less power than LIDAR, but are sensitive to changes in environmental appearance caused by factors like day-night cycles.

Radar's primary purpose in most current autonomous vehicle applications is collision avoidance: although it does not have good acuity and hence struggles to distinguish small objects, it is relatively resilient to environmental conditions such as adverse weather, and can see through smoke and fog quite well.

Finally, sensors like Global Positioning System (GPS) receivers provide positioning information (which can be disrupted by tunnels or tall

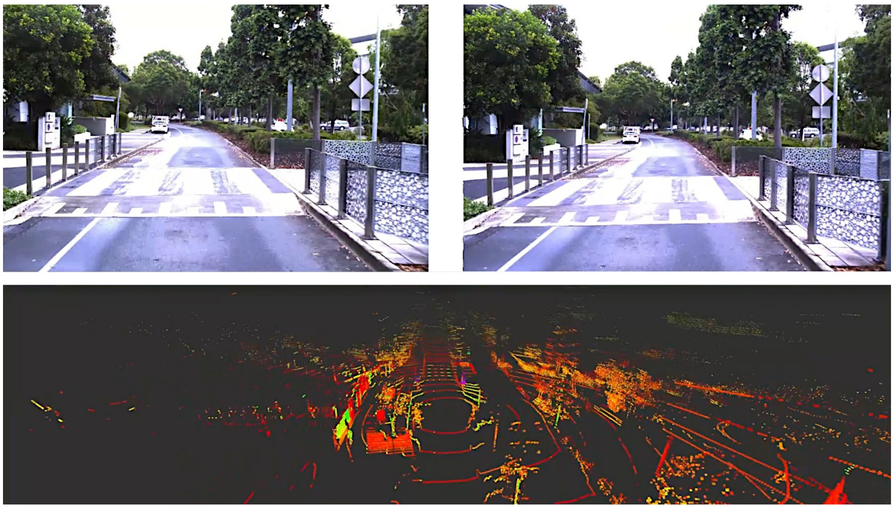

Figure 2: What an autonomous car sees: a range of camera views (top) and range scans (bottom). Photo from QUT.

buildings), while internal sensors provide information such as linear acceleration, rotational rate, steering angle and wheel speed.

2.3 Computational Hardware

Compute hardware provides the processing power to perform all the onboard autonomy-related tasks like scene understanding, navigation and high level control. To maximize electric vehicle range, recent hardware trends have focused on power usage per compute unit. Nvidia is a good example of a key player in this space, with power efficient, highly capable systems like its Jetson AGX Xavier.

Off board compute still has a useful role to play in autonomous vehicle applications, for example in the consolidation and merging of the massive amounts of data uploaded by thousands of cars in a city on a daily basis.

3 Key Technical Competencies - Software

The software operating on autonomous vehicles performs a number of key technical competencies including localization, planning, decision-making and scene understanding.

4 Mapping and Localization

Mapping and localization is a key pillar of autonomous vehicle operation. In brief, there are several subtypes of localization which play different roles in enabling autonomy on a vehicle.

Simultaneous Localization And Mapping (SLAM) has long been a major research field in robotics: how does a robot move through an environment, building up a map of that environment, whilst *simultaneously* localizing itself within that ever changing map.

Approximate localization – what you get on your phone's GPS – is typically used for overall route planning, and is obtained from GPS or from onboard localization systems. "Automation-enabling" higher precision localization is typically provided by onboard localization within existing maps of the environment, or in the case of some autonomous mining vehicles high accuracy GPS.

Relative localization is also important: for example, knowing that the vehicle is currently located 0.73 metres from the edge of the road. Accurate *relative* positioning (and velocities and accelerations) with respect to moving objects such as an oncoming car is critical for safe vehicle planning and control.

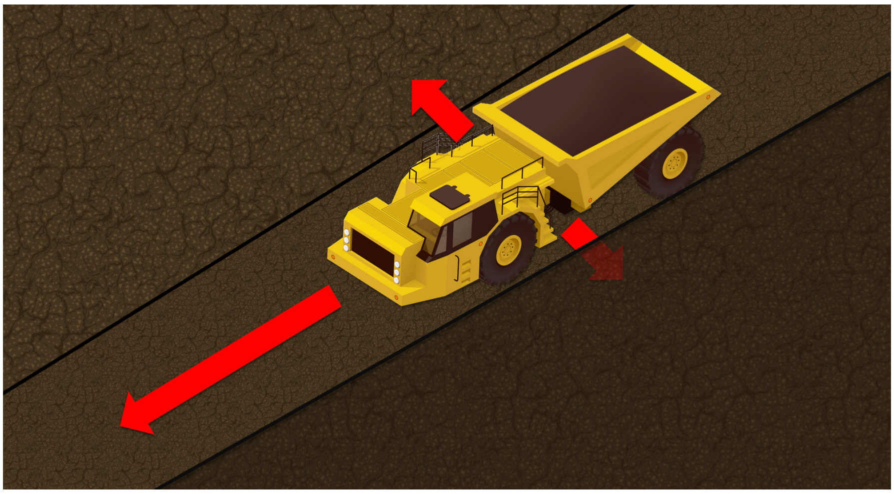

Figure 3: All errors are not created equal: for second to second control in a mining tunnel for example, minimizing lateral error is more important than downtrack (along the length of the tunnel) error, since the immediate risk is hitting the wall. Source 123rf.com / omela / tele52.

4.1 Planning, Decision-Making and Control

Just as critical to an AV's viability as sensing and mapping is what is then done with that information: how does the vehicle plan and then act, whether to accelerate, brake, turn, or activate a turning indicator. These processes play a critical role in safety: the planning system must continually plan safe actions, such as slowing down or suddenly changing lanes to avoid an unexpected obstacle when braking is not an option.

The planning and decision-making process also alters significantly when considering on-road delivery vehicles that carry goods rather than people, like Nuro's delivery vehicles. In accident situations, no longer is there any real tension between protecting humans inside and outside the vehicle: the safety of humans outside the vehicle can be entirely prioritized.

5 Interaction with Vulnerable Road Users (VRUs)

Vehicles that have reached SAE Level 3 autonomy and above will have to know how to interact with humans. This includes human drivers, who — even in a world where autonomous vehicles are rapidly adopted — will be on the roads for the foreseeable future. Bicyclists, pedestrians, motorcyclists, scooter riders: these categories of vulnerable road users have enduring claim to their share of the urban pavement. Interacting safely, explainably and politely with VRUs is likely to remain an essential part of the AV's task.

Figure 4: Detecting and predicting the intent of vulnerable road users like cyclists and pedestrians is a critical challenge for autonomous vehicles. Source Perceptive Automata.

Pedestrians and cyclists are not predictable using standard techniques like Kalman filters. Simply stopping every time a VRU could *potentially* enter the vehicle's path results in vehicles that perform excessive and unnecessary emergency maneuvers. 86% of documented incidents with AVs are either rear-endings or sideswipings that come from a human's misunderstanding of an AV's behavior. Understanding VRUs is key to eliminating this failure mode.

5.1 Moving Away from the Trolley Problem Mindset

Much of the attention devoted to interactions between autonomous vehicles and VRUs has focused on ethical dilemmas. A famous thought experiment, the "trolley problem", where a person is forced to choose which of two actions which both cause somebody's death is more morally acceptable, has been held up as a model for the kinds of decisions autonomous vehicles will have to make. While it may someday be the case that autonomous vehicles are sophisticated enough and have good enough information about the world that the primary concern with VRU interaction is how to behave ethically in the unlikely event that there is no option but to cause catastrophic bodily harm to a human, there are several reasons this is not currently a primary concern to autonomous vehicle makers.

First, the starting assumption for many vehicle makers is finding the motion plan that provably minimizes or eliminates any chance of a harm-causing interaction. The Intel division MobilEye has published work attempting to formalize risk analysis in motion planning in order to develop behavior plans where a negative interaction is impossible. Second, the types of ethical dilemmas discussed in most trolley problem research rely on very fine-grained categorization of VRUs — an old person vs. a young person, a pregnant woman vs. a helmet-less cyclist, and so on — that are largely out of reach for current perception systems in autonomous vehicles. Third, much of the current focus in AVs is minimizing harm in general, and one way to do that is to plan around the level of damage likely to be suffered. For these reasons and others, the ethical considerations raised by the trolley problem are increasingly not being considered as the most immediate practical challenge for autonomous vehicles [13].

5.2 Key Technical Breakdown

VRUs are sometimes difficult to distinguish – from the waist up, a cyclist, pedestrian and scooter rider all look highly similar to a computer vision system. Below we provide a technical breakdown of the relevant technologies that address this challenge.

5.2.1 Detecting VRUs

When we use the term detection, we mean automatically detecting that something is in the way of the vehicle. Radar, LIDAR and various other non-RGB sensors are very capable for this task, but have limitations around discriminatory resolution (radar) and current cost (LIDAR). Camera sensors represent a cheaper option that has caught on with several car companies producing self-driving vehicles.

5.2.2 Recognizing VRUs

When we use the term recognition, we mean automatically determining what exactly is in the scene as sensed by the vehicle. Current recognition algorithms are largely developed and tested in the laboratory using standard benchmark datasets like ImageNet and COCO, but a troubling disconnect exists between laboratory experimentation and real-world operation. For example, a well-known limitation of all machine learning-based algorithms is their poor handling of inputs from classes outside of the training set. This is known as the "open set" recognition problem and is a common problem in autonomous driving. A new class of open set-tolerant machine learning algorithms is being developed to address this [3].

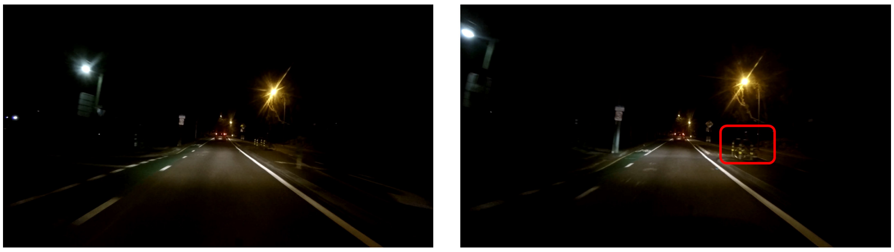

Figure 5: Reliably detecting and recognizing vulnerable road users like cyclists is difficult enough under normal conditions, but is compounded in poor visual conditions and when the VRU is partially obscured by other objects in the environment.

5.2.3 Action recognition

After detection and recognition of a VRU, there comes activity recognition - what is the VRU doing. Take one scenario: traffic officers signaling cars to follow a detour by waving in a certain direction. With a correct determination of what the officer's action means, the vehicle can alter the course it is following and safely proceed as directed.

This is a non-trivial sequence of events that must unfold within seconds and be executed with a level of accuracy that matches that of the human driver. As with other areas of visual recognition within computer vision, great strides have been made in action recognition but current approaches are not as robust as human drivers.

5.2.4 Prediction

Arguably the most important aspect of interacting with VRUs is *prediction*. A motor vehicle that is traveling straight at 25mph on a road and does not have its brake lights illuminated can be assumed to continue traveling at approximately 25mph at least momentarily. Compared with vehicles, VRUs have many fewer constraints in terms of traffic signals, other traffic, and rules of the road, and hence have much more variability in potential paths.

Much work on prediction VRUs has relied on fundamentally physics-based models; if you know the location and trajectory of the pedestrian, how well can you extrapolate their future trajectory? Elaborations have included the use of cues like the presence of relevant context like crosswalks, and the integration of information regarding the pose of the person. These approaches have proven relatively robust at very short timescales, but have not been able to provide useful predictions outside of a time window of about a second and a half. At normal urban driving speeds, that's not enough. One proposed solution is to model the dynamics of all the actors at an intersection, which critically relies on being able to accurately model every agent in the scene.

Almost all current approaches suffer from another problem, which is that the drivers that VRUs are most comfortable interacting with, humans, do nothing like either of these approaches. Humans have a finely tuned and remarkably high functioning facility called "theory of mind", which allows them to make behaviorally useful assumptions about the internal mental state of another human. A human driver isn't trying to guess the trajectory of a pedestrian; instead, they're making sophisticated inferential judgments about what that pedestrian's goals are, and how they might interact in a social process with the vehicle. Approaches which model this notion look promising.

5.2.5 Communicating car intent to VRUs

The interaction between VRUs and human-driven vehicles begins when either the driver or the VRU first notices the other, and ends when the vehicle has proceeded out of the VRU's field of view. It is bi-directional: the pedestrian wants to know that the car knows the pedestrian is there, the car wants to know what the pedestrian wants, and so on. Companies like Jaguar/Land Rover have experimented with mounting large, cartoon eyes on vehicles, to communicate information about how the AV is distributing its "attention" [6]. Former startup Drive.ai designed its vehicles to feature interactive screens which can communicate more complex messages, like "I'm waiting for you to cross". These systems have a limited grammar, but actual interactions between human drivers and VRUs also rely on a very limited grammar. To communicate with a limited grammar, the ability of both VRUs and vehicles to understand the intentions and state of mind of other road actors is essential.

6 Current Technical Issues

With the field maturing over the past fifteen years since the first DARPA Grand Challenge in 2004, it's become relatively clear that there are some key technical issues that remain unsolved, and are generally widely acknowledged by both industry and researchers working in this area. One of the most major: interaction with vulnerable road users, has already been covered. Here we briefly highlight some of the other challenges.

6.1 The Problem of Corner Cases

"Corner cases" as they have become known, are situations that rarely occur and hence are hard to predict, anticipate and react appropriately to. A person dressed in a person-sized chicken suit is one example of a corner-case. For self-driving cars the problem is particularly difficult because the current artificial intelligence techniques behind these systems do not generalize as well as a human driver, and hence have difficulty coping with these highly unusual situations. Consequently much current effort is being invested in coming up with ways to deal more effectively with these corner cases: by gathering ever larger amounts of data from the real world, by simulating billions of miles of driving, and by specifically testing pathologically difficult scenarios over and over again.

6.2 Simulation versus Real-World Testing

A key "problem" for autonomous vehicle developers is that cars are already quite safe: about one fatality for every 100 million miles of driving [11]. Consequently, it's very hard to obtain under normal conditions sufficient mileage on a limited number of development vehicles to prove the safety of a system. Developers have therefore turned to simulation as a critical tool in their autonomy arsenal. High fidelity simulation environments enable researchers to target specific weather conditions, specific pedestrian configurations, and to run much higher throughput simulation and evaluation than is possible in the real world.

A key challenge in using simulation arises from the "transferability" problem - how do you show and prove that the system you've developed in simulation will work as well in the real world, since simulation is never a perfect replication of reality. Much resource and effort is consequently invested in improving the utilization and transferability of development in simulation environments.

6.3 Sometimes versus Anytime: Weather and Other Environmental Conditions

The real-world is a constantly changing environment, which presents major challenges for autonomous vehicles. First and foremost, the environment can change in both appearance and physical structure due to day-night cycles, seasonal change, and weather conditions such as rain, snow and fog.

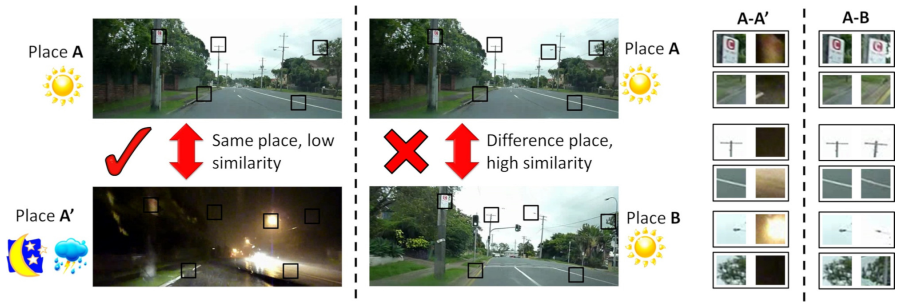

Figure 6: Environmental change poses a significant challenge for camera-based technologies: the images in the left column are from the *same* place under radically different environmental conditions, while the images in the middle column are from different places in the environment [8].

Figure 6 illustrates some of the key challenges that a changing world can cause. The same place (shown in the left column) can appear completely different at night time during a tropical storm to during clear weather in the daytime. The problem is further complicated when considering the natural environmental aliasing that can also be encountered, as shown in the middle column: these are two places that are completely different locations but look highly similar.

These problems can be partly solved using advanced methods [8] or by using sensors that are not as sensitive to appearance change like LIDAR. However, visual sensing is critical for the rich, nuanced understanding of the world around an autonomous vehicle, and consequently the problem of operating in challenging visual conditions remains a relevant, and unsolved, problem.

6.4 Provability, Explainability and Self-Characterization

A significant shortcoming of the present generation of self-driving vehicles (and deep learning in general) is the difficulty in describing the properties of their underlying deep learning models in a rigorous manner. In essence, the learning problem during training is one of function approximation, where the approximated function cannot be recovered in an exact manner afterward (this is why neural networks have a reputation of being "black boxes" [5]). We would like to be able to enforce explainability for any output of a deep learning model, but since we cannot examine any learned functions directly, we can only turn to the observable output of the system, the same situation psychologists find themselves in when studying the human brain. One possibility then is to test the deep learning models in a manner similar to how psychologists test the brain [9].

For some applications, pausing and handing off control to a human operator is feasible — but only if the system is able to assess its own performance reliably. To do this, *probabilistic outputs reflecting uncertainty* are required. For deep learning-based systems, this can be accomplished with strategies such as making small perturbations to the weights of the network, dropping out units of a trained network at test time, the use of a probabilistically calibrated readout layer or through examining statistical distributions of the data sampled by the sensors. The choice of distribution is important: *Underestimating the occurrence of rare events can be dangerous*, while *over estimating them may be problematic for usability*.

7 Autonomous Vehicles Beyond the Road

Beyond the road, autonomous vehicles are or could be deployed in a range of other domains including mining, logistics, agriculture and defence. Here we briefly cover the key deployment domains and their unique problems and opportunities.

7.1 Mining

Mining in general has several of the key characteristics that faciliated its early adoption of autonomous vehicles: it's large enough to support the capital-intensive development of autonomous vehicle-related technology, its existing remote operation workflows are more easily automated, and there are less latency-critical scenarios, meaning occasional handover to a remote operator is feasible. One example milestone in autonomous vehicles in mining: Rio Tinto's autonomous haulage system recently hauled its one billionth tonne autonomously [10].

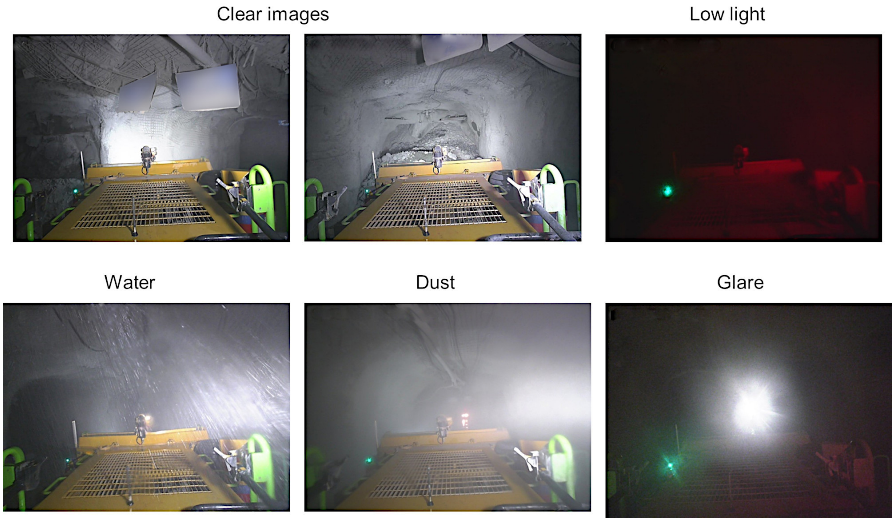

Figure 7: In underground mining environments a range of challenging perceptual conditions are encountered by autonomous vehicles including huge lighting changes, darkness, water and dust. Adapted from [14].

Mining is a challenging environment: in underground environments, there is no access to satellite-based GPS, so alternative technological solutions are required: some involve installation of additional infrastructure, local WiFi networks, or on-vehicle camera- and laser-based localization solutions. On-board camera-based solutions encounter a range of challenging perceptual conditions: dust, smoke, water, and highly varied lighting conditions. Range sensor-based solutions encounter a different set of challenges, including the highly aliased geometry of many underground tunnel systems.

7.2 Logistics

In logistics it is possible to design an entire logistics centre to facilitate higher levels of automation. Amazon's fulfillment centres, built on top of their acquisition of Kiva Systems, are a prime example of this: the autonomous robots move shelving around, rather than attempting to pick things off static shelves. Other approaches like Ocado's involve a rigid square lattice upon which robots move around, picking up and dropping off grocery loads. In both cases, humans are restricted to certain areas of the environment, so human safety concerns are significantly reduced as a technological concern.

7.3 Agriculture

Figure 8: Agriculture shares many of the same motivations for autonomous vehicles as mining, but widespread commercial deployment has lagged [2].

Farms are generally areas with relatively controlled access and minimal to no human presence in the operational zone of an autonomous vehicle. They are also sometimes areas in which human labour can be hard to find, further motivating the case for developing autonomous vehicles. Autonomous farming vehicles can perform a range of activities, including sowing and planting crops, killing weeds and the long term holy grail: harvesting crops. Progress has been slow: while there have been dozens of autonomous vehicle trials, there are few long term commercial deployments. Most of the more capable platform demonstrations have only been announced in the past 2-3 years [7].

7.4 Defence

In defence, like in mining, the cost per unit of many vehicle types is typically far larger than a normal consumer car, enabling the use of more capable sensing and compute. Much modern defence theory assumes that there will be a complete blackout on both communications and GPS-based positioning technologies (similar to the conditions imposed on underground autonomous mining trucks): meaning on-vehicle autonomy will have to shoulder the bulk of the decision-making itself, rather than rely on outsourcing to a human at a remote command post.

The environments that these vehicles might deploy into, like ruined, dusty or smoking urban landscapes and thickly vegetated forests pose a range of challenging mobility, perception, planning and control challenges. Finally there are also the ethical considerations around autonomy in any defence application as well: one that is receiving ongoing and significant sustained attention [4].

7.5 Other Fields in Brief

There are almost 40 marine ports that are at least partly automated globally [12], with some of those autonomous components involving autonomous vehicles, for example shifting shipping containers around. Other areas of autonomous vehicle deployment include sidewalk-based delivery vehicles like Amazon's Scout program and Starship technologies. These vehicles are typically relatively small, cheap and move at relatively low speeds, radically reducing their danger profile compared to on-road larger vehicles moving at higher speeds.

8 Conclusion

Autonomous vehicle-enabling technology has matured and advanced significantly over the past decade in a range of domains including on-road passenger-carrying or delivery vehicles, mining and logistics. In some application areas such as logistics and mining these vehicles already form a commercially critical part of the companies that operate them, while in others, most notably on-road autonomous vehicles, widespread commercial deployment has still not occurred.

Much of the core technology is likely to continue benefitting from steady progress in sensing and compute capability (along with a corresponding decrease in price), and the associated progress in vital technical capabilities like general scene understanding and vulnerable road user interaction. In fields where safety is not directly involved, such as those where humans are physically absent from the operating environment of autonomous vehicles, future progress will likely be determined by simple commercial calculations based on the cost and efficiency of autonomous vehicle systems. But for widespread on-road deployment, there remains key technical hurdles to overcome and demonstrate with respect to safety, which makes for interesting years ahead.

9 Authors

9.1 Michael Milford

Michael Milford is a Professor at the Queensland University of Technology and a Chief Investigator at the Australian Centre for Robotic Vision. His

research models the neural mechanisms in the brain underlying tasks like navigation and perception to develop new technologies in challenging application domains such as all-weather, anytime positioning for autonomous vehicles.

9.2 Sam Anthony

Sam Anthony is an ex-hacker, vision scientist, expert in human cognition, and veteran of tech booms and busts. He is the CTO and Co-Founder of Perceptive Automata, Inc., an early-stage company building safety systems for autonomous vehicles.

9.3 Walter Scheirer

Walter Scheirer is an Assistant Professor in the Department of Computer Science and Engineering at the University of Notre Dame, and a Co-Founder of Perceptive Automata, Inc. His overarching research interest is the fundamental problem of recognition, including the representations and algorithms supporting solutions to it.

References

[1] *$80 billion has been spent on self-driving cars with nothing to show for it. | TreeHugger*. https://www.treehugger.com/cars/praise-dumb-transportation-self-driving.html. (Accessed on 06/12/2019).

[2] Owen Bawden et al. "Robot for weed species plant-specific management". In: *Journal of Field Robotics* 34.6 (2017), pp. 1179–1199.

[3] Terrance Boult et al. "Learning and the Unknown: Surveying Steps Toward Open World Recognition". In: *AAAI Conference on Artificial Intelligence (AAAI 2019)*. https://www.vast.uccs.edu/~tboult/PAPERS/Learning_and_the_Unknown_Surveying_Steps_Toward_Open_World_Recognition_AAAI19.pdf. 2019.

[4] *Frontiers | A Review of Future and Ethical Perspectives of Robotics and AI | Robotics and AI*. https://www.frontiersin.org/articles/10.3389/frobt.2017.00075/full. (Accessed on 06/12/2019).

[5] *Inside the 'Black Box' of a Neural Network | WIRED*. https://www.wired.com/story/inside-black-box-of-neural-network/. (Accessed on 06/12/2019).

[6] *Jaguar Land Rover is putting googly eyes on its autonomous cars*. https://www.fastcompany.com/90231563/people-dont-trust-autonomous-vehicles-so-jaguar-is-adding-googly-eyes. (Accessed on 06/12/2019).

[7] *John Deere develops fully electric, autonomous tractor | Industrial Vehicle Technology International*. https://www.ivtinternational.com/news/agriculture/john-deere-develops-fully-electric-autonomous-tractor.html. (Accessed on 06/12/2019).

[8] M. J. Milford and G. F. Wyeth. "SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights". In: *2012 IEEE International Conference on Robotics and Automation*. May 2012, pp. 1643–1649. DOI: 10.1109/ICRA.2012.6224623.

[9] Brandon RichardWebster, Samuel E. Anthony, and Walter J. Scheirer. "PsyPhy: A Psychophysics Driven Evaluation Framework for Visual Recognition". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)* (2019). https://ieeexplore.ieee.org/document/8395028.

[10] *Rio Tinto's autonomous trucks pass one billion tonne milestone - Australian Mining*. https://www.australianmining.com.au/news/rio-tintos-autonomous-trucks-pass-one-billion-tonne-milestone/. (Accessed on 06/05/2019).

[11] *State by state*. https://www.iihs.org/topics/fatality-statistics/detail/state-by-state. (Accessed on 06/07/2019).

[12] *The future of port automation | McKinsey*. https://www.mckinsey.com/industries/travel-transport-and-logistics/our-insights/the-future-of-automated-ports. (Accessed on 06/05/2019).

[13] *Why the Trolley Dilemma for safe self-driving cars is flawed*. https://www.fastcompany.com/90308968/why-the-trolley-dilemma-is-a-terrible-model-for-trying-to-make-self-driving-cars-safer. (Accessed on 06/12/2019).

[14] Fan Zeng et al. "Enhancing underground visual place recognition with Shannon Entropy saliency". In: (2017).