ECAI 2020
G.D. Giacomo et al. (Eds.)

doi:10.3233/FAIA200463

Driven by Commonsense

On the Role of Human-Centred Visual Explainability for Autonomous Vehicles

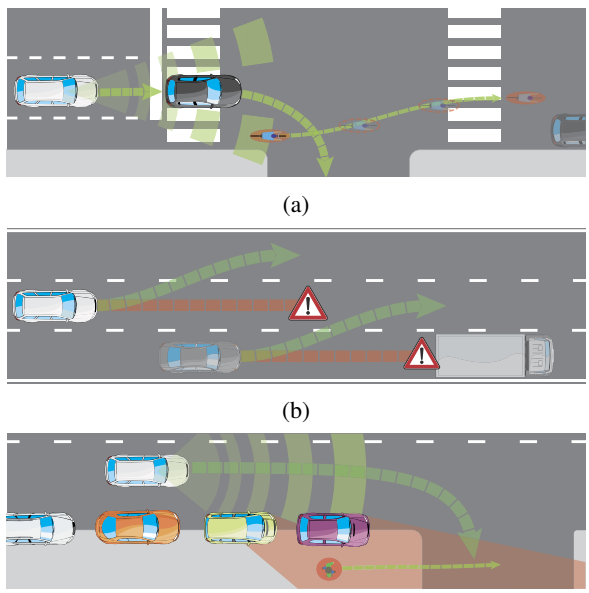
Jakob Suchan[1] and **Mehul Bhatt**[2] and **Srikrishna Varadarajan**
CoDesign Lab / Cognitive Vision – www.codesign-lab.org/cognitive-vision

Abstract. Within the autonomous driving domain, there is now a clear need and tremendous potential for hybrid solutions (e.g., integrating semantics, learning, visual computing) towards fulfilling essential legal and ethical responsibilities involving explainability (e.g., for diagnosis), human-centred AI (e.g., interaction design), and industrial standardisation (e.g, pertaining to representation, realisation of rules & norms). In these contexts, this highlight paper positions recent research from IJCAI 2019 [4] aimed at advancing human-centred AI principles in the backdrop of the autonomous driving application domain. From a technical viewpoint, the highlighted research provides a model for advancing the state of the art in reasoning about space and motion, combining reasoning and learning, non-monotonic reasoning, and computational modelling of high-level visuospatial commonsense. In addition to demonstrating the significance of integrated vision and semantics solutions in autonomous driving, we also highlight open questions emphasising the need for interdisciplinary mixed-methods research –involving AI, Psychology, HCI– to better appreciate the complexity and spectrum of varied human-centred challenges in diverse naturalistic driving situations.

1 Human-Centred Autonomous Driving

The Federal Ministry of Transport and Digital Infrastructure in Germany (BMVI) has taken a lead in eliciting 20 key propositions (with possible legal implications) for the fulfilment of ethical commitments for automated and connected driving systems [2]. The BMVI report highlights a range of factors pertaining to safety, utilitarian considerations, human rights, statutory liability, technological transparency, data management and privacy etc. We claim that what appears as spectrum of complex challenges (in autonomous driving) that may possibly delay technology adoption is actually rooted to one fundamental methodological consideration that needs to be prioritised, namely: the design and implementation of human-centred technology based on a *confluence* of techniques and perspectives from AI+ML, Cognitive Science & Psychology, Human-Machine Interaction, and Design Science. Such an integrative approach has so far not been explored within autonomous driving research.

Driving / Standardisation. Regulation. The recent surge in autonomous vehicle research has coincided with (and been driven by) advances in *deep learning* based computer vision research. Although deep learning based vision & control have (arguably) been successful for self-driving vehicles, we posit that there is a clear need and tremendous potential for hybrid visual sensemaking solutions integrating *vision and semantics* towards fulfilling essential legal and ethical responsibilities (such as those in [2]) involving explainability, human-centred AI, and industrial standardisation (e.g, pertaining to representation, realisation of rules and norms). As the autonomous driving vehicle industry develops, it will be necessary —e.g., similar to sectors such as medical computing, computer aided design— to have an articulation and community consensus on aspects such as representation, interoperability, human-centred performance benchmarks, and data archival & retrieval mechanisms. In spite of major investments in self-driving vehicle research, issues related to human-centred'ness, human collaboration, and standardisation have been barely addressed, with the current focus in driving research primarily being on two basic considerations: *how fast to drive, and which way and how much to steer*. This is necessary, but inadequate if autonomous vehicles are to become commonplace and function with humans. Ethically driven standardisation and regulation will require addressing challenges in semantic visual interpretation, natural / multimodal human-machine interaction, high-level data analytics (e.g., for post hoc diagnostics, dispute settlement) etc. This will necessitate –amongst other things– human-centred qualitative benchmarks and integration of multifaceted hybrid AI solutions.



(a)

(b)

(c)

Figure 1: Safety-Critical Situations (select sample): **(a)**. momentarily occluded / hidden entities; **(b)**. overtaking / lane-crossing situation; and **(c)**. blocked visibility.

[1] University of Bremen, GERMANY
[2] Örebro University, SWEDEN

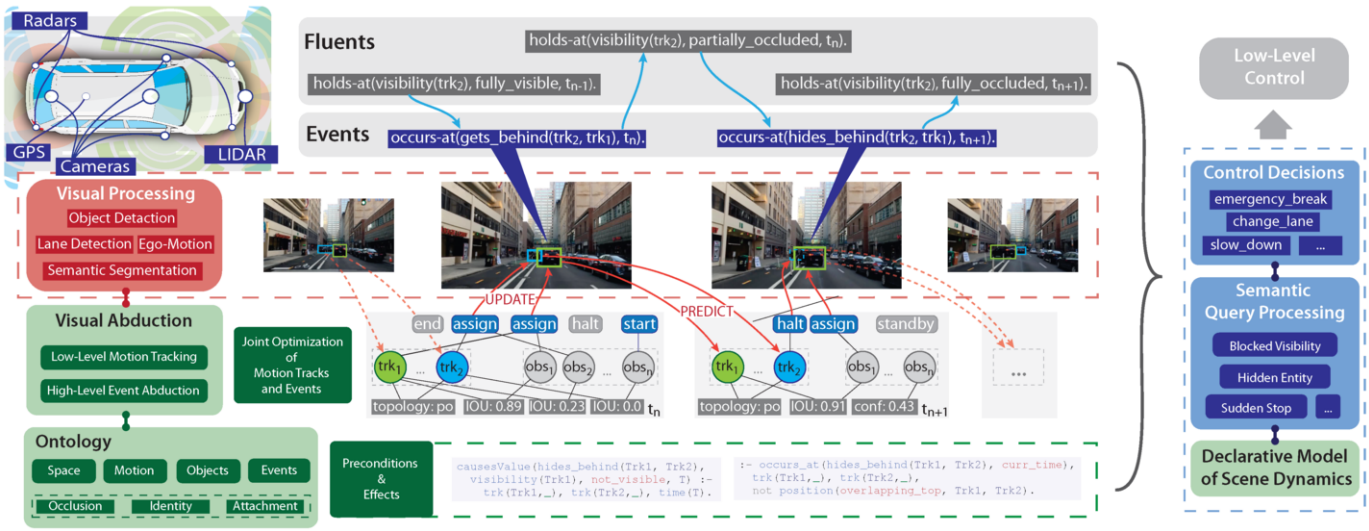


Figure 2: Commonsense Visual Explainability in Active Vision & Control (for Autonomous Driving); The Case of Hidden Entities.

2 Driven by Visual Commonsense

The work [4] positioned in this highlight paper is motivated by human-centred considerations in the design and implementation of (visual sensemaking) technology for autonomous vehicles. We aim to demonstrate the significance of semantically-driven methods rooted in knowledge representation and reasoning (KR) in addressing research questions pertaining to explainability and human-centred AI particularly from the viewpoint of sensemaking of dynamic visual imagery. In particular, with Answer Set Programming (ASP) as an underlying knowledge representation language and robust methodology for non-monotonic reasoning [3], our work (Fig. 2) bridges the gap between high-level formalisms for logical abduction and low-level visual processing by tightly integrating semantic abstractions of space-time-motion with their underlying numerical representations in the backdrop of key considerations pertaining to semantics, commonsense reasoning, and explainability [1].

Safety-Critical Commonsense (An Example). Figure 1 schematises select –seemingly common but computationally challenging– safety-critical situations; lets take case of *occlusion* (Fig. 1a):

Car (c) is in-front, and indicating to turn-right; during this time, person (p) is on a bicycle (b) and positioned front-right of c and moving-forward. Car c turns-right, during which the bicyclist $< p, b >$ is not visible. Subsequently, bicyclist $< p, b >$ reappears.

The occlusion scenario alone indicates several challenges concerning aspects such as: identity maintenance of occluded objects, making default assumptions about inertia, computing "what-if" counterfactuals aimed at the cognitive equivalent of mental simulations, temporal projection, and interpolation of missing information (e.g., what could be hypothesised about bicyclist $< p, b >$ when it is *occluded*; how can this hypothesis enable in planning an immediate next step). Addressing such challenges —be it realtime or post-hoc— in view of human-centred AI concerns pertaining to ethics, human-centred explainability and regulation requires a systematic integration of **Semantics and Vision**, i.e., robust commonsense representation & inference about *space, actions, and change* on the one hand, and powerful low-level visual computing capabilities, e.g., pertaining to object detection and tracking on the other. The highlighted work of [4] presents a computational model its evaluation with active vision and occlusion problems in a realtime setting.

3 An Open Road

Our development of a systematic, modular, and general visual sensemaking methodology in [4] opens up several possibilities for extensions: specialised commonsense theories about multi-sensory integration, multi-agent belief merging, incorporation of contextual knowledge and situational norms etc based on requirements, experimenting with specialised visual computing methods. Furthermore, not everything about autonomous vehicles is about realtime control / decision-making; several human-machine interaction requirements (e.g., for diagnostic communication, universal design) also exist. Our own ongoing focus is to develop a novel dataset emphasising semantics and (commonsense) explainability; this is driven by mixed-methods research –AI, Psychology, HCI– for the study of driving behaviour in low-speed, complex urban environments with unstructured traffic. Here, emphasis is on natural interactions (e.g., gestures, joint attention, visual search complexity) amongst drivers, pedestrians, cyclists etc. Such interdisciplinary studies are needed to better appreciate the complexity and spectrum of varied human-centred challenges in autonomous driving, and demonstrate the significance of integrated vision & semantics solutions in those contexts.

Acknowledgement. Partial funding by the German Research Foundation (DFG) via the CRC 1320 EASE – "Everyday Activity Science and Engineering", Project P3: Spatial Reasoning in Everyday Activity is acknowledged.

REFERENCES

[1] Mehul Bhatt, 'Reasoning about space, actions and change: A paradigm for applications of spatial reasoning', in *Qualitative Spatial Representation and Reasoning: Trends and Future Directions*. IGI Global, (2012).

[2] BMVI. Report by the Ethics Commission on Automated and Connected Driving., BMVI: Federal Ministry of Transport and Digital Infrastructure, Germany, 2018.

[3] Gerhard Brewka, Thomas Eiter, and Miroslaw Truszczyński, 'Answer set programming at a glance', *Commun. ACM*, **54**(12), 92–103, (2011).

[4] Jakob Suchan, Mehul Bhatt, and Srikrishna Varadarajan, 'Out of Sight But Not Out of Mind: An Answer Set Programming Based Online Abduction Framework for Visual Sensemaking in Autonomous Driving', in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, ed., Sarit Kraus, pp. 1879–1885. ijcai.org, (2019).