Social Vision for Intelligent Vehicles: From Computer Vision to Foundation Vision

Hui Yu, *Senior Member, IEEE*, Yutong Wang, Yonglin Tian, *Member, IEEE*, Hui Zhang, Wenbo Zheng, *Member, IEEE*, Fei-Yue Wang, *Fellow, IEEE*

***Abstract*—The development of Artificial intelligent (AI), deep learning and computer vision technologies has enabled various new applications and features of intelligent vehicles. However, there are still application scenarios that have not been addressed, such as the new social and human-vehicle interaction within the intelligent vehicle cabin, where social vision could play an important role. To this end, this letter presents a summary of discussions as part of Distributed/Decentralized Hybrid Workshop on Sustainability for Transportation and Logistics (DHW-STL) dedicating to the development of sustainable interaction systems.**

***Index Terms*—Social Vision, Parallel Vision, Knowledge Vision, Foundation Vision, intelligent vehicles, social interaction, sustainability**

I. Introduction

Artificial intelligent (AI) and computer vision technologies have been widely used to capture and analyze social signals, human movements, and emotions to assist interaction between humans and machines [1]–[8]. For example, body gestures can express specific behavioral intentions [9], which are widely analyzed to interact with robots and computers. Facial expressions convey rich emotional information and thus play an important role in social and human-machine interaction. Jack et al. [10] even discovered significant cultural differences in emotional facial expression between western and eastern people, which provided a further understanding of emotional expression for social and human-machine interaction.

Recently, research suggested that the popular large language model ChatGPT [11]–[14] could be brought to the field of intelligent vehicles to improve interaction experience. However, there is still a range of unexplored scenarios for social and human-machine interaction with the prevalence of intelligent vehicles, such as smart vehicle cabins, driverless public buses, training, and entertainment venues.

The advent of intelligent vehicles will enable significant changes in the passengers' role and behaviour without the need of actually driving the vehicles. The interaction among the passengers and even between passengers and the vehicle will also have new and dramatic changes resulting in a shift in behaviours. Furthermore, compared with the conventional model where drivers directly control vehicles, passengers may have more direct interaction with the intelligent vehicles and other passengers. In such scenarios, the question of how passengers should engage in social interactions during the journey and communicate with the vehicle becomes a crucial issue. These challenges require fundamental changes in the design and application of new related technologies. In this letter, we summarize the discussions held at the Distributed/Decentralized Hybrid Workshop on Sustainability for Transportation and Logistics (DHW-STL) about Social Vision for intelligent vehicles and their relationships in the context of autonomous driving and social interaction. The main concepts discussed include Foundation Vision, Knowledge Vision, Parallel Vision and Social Vision.

II. Social Vision for Intelligent Vehicles

We first present four key concepts: Foundation Vision, Knowledge Vision, Social Vision and Parallel Vision. Then, we briefly report the discussion on their relationships.

Foundation Vision refers to a new vision framework consisting of a vision operation system based on the foundation model and active vision with the ability to handle versatile foundational large models, multimodal data and scenario models. It has a comprehensive capability to deal with various vision-based tasks, such as object detection [15], image segmentation [16], emotional expression [17], reconstruction [18], and classification [19]. Foundation Vision provides an effective way to augment humans and machines as well as automating the processing of large models and scenarios from a diverse range of sensors or data sources, including images, text, sound, cognitive signals, etc.

Knowledge Vision draws inspiration from the field of knowledge engineering, aiming primarily to incorporate human knowledge into computer vision models [20]–[22]. Usually, a structured methodology is employed to represent human knowledge and establish structured knowledge that directs the training of visual models. By following this approach, visual models can acquire improved perceptual and understanding capabilities, as discussed in [21].

Parallel Vision is a systematic virtual-real framework, which is derived from the artificial societies, computational experiments, and parallel execution (ACP) technology in parallel control systems [23]–[26]. It shares the fundamental concept with the digital twin [27]–[30] but with a more comprehensive

Manuscript received April 19, 2023; revised August 16, 2023.

H. Yu is with the School of Creative Technologies, University of Portsmouth, Portsmouth, PO1 2DJ, UK, Email: hui.yu@port.ac.uk

Y. Wang, Y. Tian and F.-Y. Wang are with the State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, 100190, China

W. Zheng is with School of Computer Science and Artificial Intelligence, Wuhan University of Technology, Wuhan 430070, China, and is also with the Sanya Science and Education Innovation Park of Wuhan University of Technology, Sanya 572000, China.

H. Zhang is with Beijing Jiaotong University, Beijing, China

Corresponding author: Fei-Yue Wang, Email: feiyue.wang@ia.ac.cn

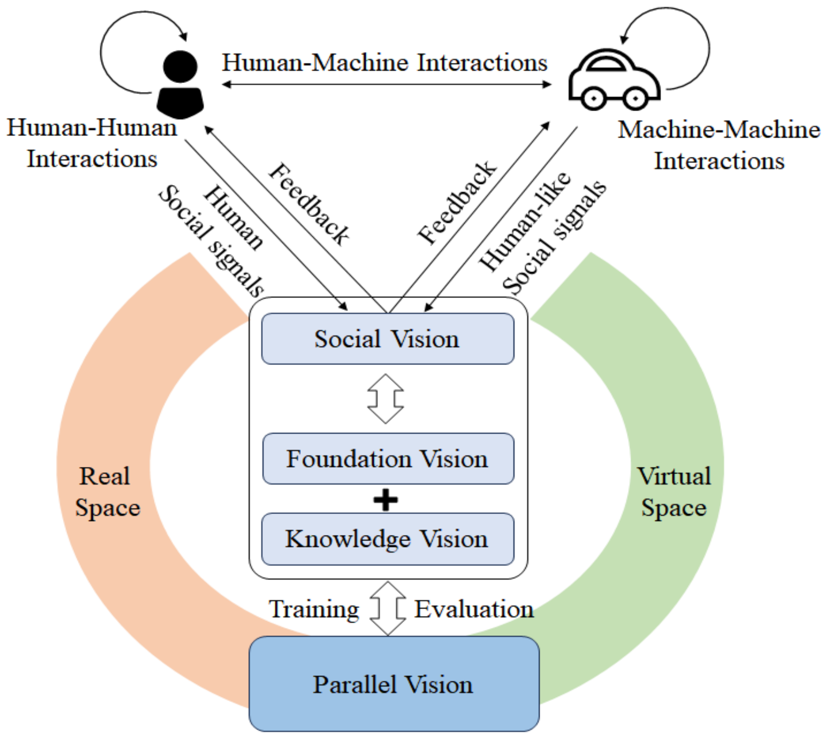

Fig. 1. The overall architecture of Social Vision for intelligent vehicles

and theoretical elaboration. Parallel Vision includes the parallel operation of virtual simulation systems and real systems. This involves running artificial scenarios in a virtual simulation to explore possible real-world occurrences and using parallel execution to guide and optimize the real systems [31].

Social Vision can be traced back to its origin in social psychology within the realm of social sciences and visual research in natural sciences [32], [33]. Social Vision plays a critical role in the development and maintenance of social cognition, which relies on the human visual system to process visual cues during interaction. In [34], Wang suggested a new direction and possibility combining social vision in psychology with computer vision in engineering and science that redefined social vision as an interdisciplinary field for intelligent science, technology, and social studies. In this letter, we extend social vision to a broader concept underpinned by foundation vision focusing the application on interaction in terms of intelligent vehicles. It includes the process of obtaining human's (e.g. passengers) emotions, physical behaviors, and intentions from their social signals, such as body gesture, facial expressions, speech, heart rate, and blood pressure. Subsequently, these signals are integrated into Foundation Vision and Knowledge Vision, through establishing a comprehensive framework to facilitate concurrent operation in both the virtual and real domains using Parallel Vision.

Fig. 1 shows the overall architecture of the Social Vision for intelligent vehicles. During the operation of intelligent vehicles, signals are captured, perceived and processed through the communication of Foundation Vision, Social Vision and Knowledge Vision. Parallel Vision provides an essential platform for interactive training and evaluation of the data from these three spaces.

III. A Vision for Transportation 5.0 and Industry 5.0

To conclude, this letter presents a Social Vision framework based on the report [34] and the DHW-STL workshop discussion, specifically for the intelligent vehicles application. This framework provides the first step toward a vision for the social and human-vehicle interaction in the context of intelligent vehicles [35], [36]. With the prevalence of metaverse technologies [37], [38], the integration of Social Vision with the metaverse will open a door for a broader spectrum of social and human-machine interactions, which would play an important role in Transportation 5.0 and Industry 5.0 [39]–[41].

References

[1] B. Tian, Y. Li, B. Li, and D. Wen, "Rear-view vehicle detection and tracking by combining multiple parts for complex urban surveillance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 2, pp. 597–606, 2014.

[2] L. Li *et al.*, "Artificial intelligence test: A case study of intelligent vehicles," *Artificial Intelligence Review*, vol. 50, pp. 441–465, 2018.

[3] S. Fan *et al.*, "Scf-net: Learning spatial contextual features for large-scale point cloud segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 14 504–14 513.

[4] W. Zhou, S. Dong, J. Lei, and L. Yu, "Mtanet: Multitask-aware network with hierarchical multimodal fusion for rgb-t urban scene understanding," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 1, pp. 48–58, 2022.

[5] G. Chen, F. Wang, X. Yuan, Z. Li, Z. Liang, and A. Knoll, "Neurobiometric: an eye blink based biometric authentication system using an event-based neuromorphic vision sensor," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 1, pp. 206–218, 2020.

[6] T. Zhou, J. Chen, Y. Shi, K. Jiang, M. Yang, and D. Yang, "Bridging the view disparity between radar and camera features for multi-modal fusion 3d object detection," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 2, pp. 1523–1535, 2023.

[7] H. Mo, Y. Meng *et al.*, "Interval type-2 fuzzy hierarchical adaptive cruise following-control for intelligent vehicles," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 9, pp. 1658–1672, 2022.

[8] Y. Xia, W. Zheng, Y. Wang, H. Yu, J. Dong, and F.-Y. Wang, "Local and global perception generative adversarial network for facial expression synthesis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 3, pp. 1443–1452, 2022.

[9] C. Zhu, J. Yang, Z. Shao, and C. Liu, "Vision based hand gesture recognition using 3D shape context," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 9, pp. 1600–1613, 2019.

[10] R. E. Jack, O. G. Garrod, H. Yu, R. Caldara, and P. G. Schyns, "Facial expressions of emotion are not culturally universal," *Proceedings of the National Academy of Sciences*, vol. 109, no. 19, pp. 7241–7244, 2012.

[11] F.-Y. Wang, Q. Miao, X. Li, X. Wang, and Y. Lin, "What does chatgpt say: The dao from algorithmic intelligence to linguistic intelligence," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 3, pp. 575–579, 2023.

[12] Y. Tian *et al.*, "VistaGPT: Generative parallel transformers for vehicles with intelligent systems for transport automation," *IEEE Transactions on Intelligent Vehicles*, 2023.

[13] H. Du, S. Teng, H. Chen, J. Ma, X. Wang, C. Gou, B. Li, S. Ma, Q. Miao, X. Na *et al.*, "Chat with chatgpt on intelligent vehicles: An ieee tiv perspective," *IEEE Transactions on Intelligent Vehicles*, 2023.

[14] Y. Wang, X. Wang, X. Wang, J. Yang, O. Kwan, L. Li, and F.-Y. Wang, "The chatgpt after: Building knowledge factories for knowledge workers with knowledge automation," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 11, pp. 2041–2044, 2023.

[15] J. Wang *et al.*, "Parallel vision for long-tail regularization: Initial results from ivfc autonomous driving testing," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 2, pp. 286–299, 2022.

[16] W. Zhou, S. Dong, J. Lei, and L. Yu, "Mtanet: Multitask-aware network with hierarchical multimodal fusion for rgb-t urban scene understanding," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 1, pp. 48–58, 2022.

[17] H. Yu, O. G. Garrod, and P. G. Schyns, "Perception-driven facial expression synthesis," *Computers & Graphics*, vol. 36, no. 3, pp. 152–162, 2012.

[18] J. Gao, D. Kong, S. Wang, J. Li, and B. Yin, "Cignet: Category-and-intrinsic-geometry guided network for 3D coarse-to-fine reconstruction," *Neurocomputing*, vol. 554, p. 126607, 2023.

[19] N. Yu, R. Yang, and M. Huang, "Deep common spatial pattern based motor imagery classification with improved objective function," *International Journal of Network Dynamics and Intelligence*, pp. 73–84, 2022.

[20] W. Zheng *et al.*, "Computational knowledge vision: paradigmatic knowledge based prescriptive learning and reasoning for perception and vision," *Artificial Intelligence Review*, vol. 55, no. 8, pp. 5917–5952, 2022.

[21] F.-Y. Wang, "Computing knowledge vision: Toward new paradigm for intelligent visual reasonings," *CAAI Transactions on Intelligent Systems*, vol. 16, no. 3, pp. 1–1, 2021.

[22] G. Zhao, Y. Li, and Q. Xu, "From emotion ai to cognitive ai," *International Journal of Network Dynamics and Intelligence*, pp. 65–72, 2022.

[23] K. Wang *et al.*, "Parallel vision: An acp-based approach to intelligent vision computing," *Acta Automatica Sinica*, vol. 42, no. 10, pp. 1490–1500, 2016.

[24] Wang, Fei-Yue, "Parallel control and management for intelligent transportation systems: Concepts, architectures, and applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 3, pp. 630–638, SEP 2010.

[25] F.-Y. Wang, "The dao to metacontrol for metasystems in metaverses: The system of parallel control systems for knowledge automation and control intelligence in cpss," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 11, pp. 1899–1908, 2022.

[26] F.-Y. Wang, "New control paradigm for industry 5.0: From big models to foundation control and management," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 8, pp. 1643–1646, 2023.

[27] Z. Wang *et al.*, "A new era of intelligent vehicles and intelligent transportation systems: Digital twins and parallel intelligence," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 4, pp. 2619 – 2627, 2023.

[28] Z. Hu, S. Lou, Y. Xing, X. Wang, D. Cao, and C. Lv, "Review and perspectives on driver digital twin and its enabling technologies for intelligent vehicles," *IEEE Transactions on Intelligent Vehicles*, 2022.

[29] P. Ye and F.-Y. Wang, *Parallel Population and Parallel Human*, 1st ed. Wiley-IEEE Press, June 2023.

[30] P. Ye *et al.*, "Parallel population and parallel human—a cyber-physical social approach," *IEEE Intelligent Systems*, vol. 37, no. 5, pp. 19–27, 2022.

[31] L. Chen *et al.*, "Parallel driving os: A ubiquitous operating system for autonomous driving in cpss," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 4, pp. 886–895, 2022.

[32] R. B. Adams, *The science of social vision: The science of social vision*. Oxford University Press, 2011, vol. 7.

[33] R. B. Adams Jr, D. N. Albohn, and K. Kveraga, "Social vision: Applying a social-functional approach to face and expression perception," *Current directions in psychological science*, vol. 26, no. 3, pp. 243–248, 2017.

[34] F.-Y. Wang, "Social vision for parallel vision in cpss: A new perspective for social computing and visual reasoning," *QAII Technical Report*, August 2020.

[35] Y. Wang, M. Kang, Y. Liu, J. Li, K. Xue, X. Wang, J. Du, Y. Tian, Q. Ni, and F.-Y. Wang, "Can digital intelligence and cyber-physical-social systems achieve global food security and sustainability?" *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 11, pp. 2070–2080, 2023.

[36] S. Teng, X. Hu, P. Deng, B. Li, Y. Li, Y. Ai, D. Yang, L. Li, Z. Xuanyuan, F. Zhu *et al.*, "Motion planning for autonomous driving: The state of the art and future perspectives," *IEEE Transactions on Intelligent Vehicles*, 2023.

[37] Q. Miao *et al.*, "Dao to hanoi via desci: Ai paradigm shifts from alphago to chatgpt," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 4, pp. 877–897, 2023.

[38] F.-Y. Wang, "Parallel intelligence in metaverses: Welcome to hanoi!" *IEEE Intelligent Systems*, vol. 37, no. 1, pp. 16–20, 2022.

[39] F.-Y. Wang, Y. Lin, P. A. Ioannou, L. Vlacic, X. Liu, A. Eskandarian, Y. Lv, X. Na, D. Cebon, J. Ma, L. Li, and C. Olaverri-Monreal, "Transportation 5.0: The dao to safe, secure, and sustainable intelligent transportation systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 10, pp. 10 262–10 278, 2023.

[40] F.-Y. Wang, J. Yang, X. Wang, J. Li, and Q.-L. Han, "Chat with chatgpt on industry 5.0: Learning and decision-making for intelligent industries," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 4, pp. 831–834, 2023.

[41] X. Wang, J. Yang, Y. Wang, Q. Miao, F.-Y. Wang, A. Zhao, J.-L. Deng, L. Li, X. Na, and L. Vlacic, "Steps toward industry 5.0: Building "6s" parallel industries with cyber-physical-social intelligence," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 8, pp. 1692–1703, 2023.