Application of Multi-Modal Interaction Based on Augmented Reality to Enhance the Environment Perception and Cabin Experience of Intelligent Car

Gangqi Wu[a]
Maynooth International School of Engineering, Fuzhou University, Fuzhou, Fujian, China

Keywords: Intelligent Vehicles, Multi-Modal Interaction, AR.

Abstract: The application of human-computer interaction in intelligent vehicles has become quite mature, which improved the performance of the vehicle. However, the use of multi-modal interactions presents redundancy, and the resulting functional stacking and conflict issues urgently need to be resolved. Therefore, the fusion strategy of multi-modal interaction in intelligent vehicles is particularly important. In order to further upgrade intelligent vehicles, this study analyzed the limitations of current interaction methods on deficiencies from the perspectives of vehicle environment perception ability and cabin experience. Then, starting from augmented reality, the research integrates the interaction modes of various modalities through a car mounted AR system, alleviating the limitations of single modal interaction and the problems caused by rigid fusion of multi-modal interaction. This can significantly enhance the safety, convenience, and intelligence of vehicles. This paper aims to provide theoretical support and practical ideas for the multi-modal interaction fusion of intelligent vehicle systems, and to provide a reference for more efficient and user-friendly human-computer interaction design in the future.

1 INTRODUCTION

Modern intelligent vehicles are widely used and developing rapidly, which urgently needs to address issues such as safety, practicality, and comfort. Compared to the traditional view of "people adapting to cars" in automobiles, intelligent cars aim to "adapt to people" and gradually transition towards "mutual adaptation between people and cars"(Gao et al., 2024). Therefore, in order to address the multifaceted issues in the development of intelligent vehicles, researchers need to consider both the "human" and "environmental" aspects, not only focusing on improving the driver's cabin experience but also taking into account the enhancement of the vehicle's environmental perception ability. This study focuses on the application of multi-modal interaction in intelligent vehicles, for comprehensively promoting the upgrading and development of intelligent vehicles in terms of environmental perception and cabin experience.

The early field of intelligent cars focused on "autonomous driving" and conducted relatively simple research on remote-controlled cars. It was not until 1969 that John McCarthy, the father of artificial intelligence, proposed in his book "Computer Controlled Cars" that intelligent cars shifted from "human control" to "computer control" when users input a destination to drive the car. But the research scope of intelligent cars still remains at the stage of intelligent driving, without considering the driver's experience in the vehicle cabin. Low level human-computer interaction still cannot make it a truly intelligent car. In recent years, significant progress has been made in the research of intelligent vehicles. Researchers are focusing on building models and improving the Intelligent Driver Model (IDM). There are models in vehicle environment perception, such as modeling the following behavior and driving style of intelligent cars based on generative adversarial learning, and upgrading the longitudinal following system (Zhou et al., 2020; Xu et al., 2024). At the same time, the application of various interactive

[a] https://orcid.org/0009-0005-5058-3659

Wu, G.
Application of Multi-Modal Interaction Based on Augmented Reality to Enhance the Environment Perception and Cabin Experience of Intelligent Car.
DOI: 10.5220/0014355400004718

In *Proceedings of the 2nd International Conference on Engineering Management, Information Technology and Intelligence (EMITI 2025)*, pages 359-364
ISBN: 978-989-758-792-4

methods in intelligent cockpits, such as using electroencephalography(EEG) signals to observe the driver's cognitive response and examining the driver's color perception based on visual neural networks (Zhang et al., 2024; Zhao, 2022), is used to enhance the user's cockpit experience. Research on modeling and analysis of multi-modal interaction methods both inside and outside the cabin, continuously improving the research system of intelligent vehicles.

However, the existing studies on intelligent vehicles often focus on the perception and recognition ability of vehicles, specifically analyzing the application of core technologies such as artificial intelligence, edge computing, sensors and other technologies in vehicles, or simply examining the human-machine interaction level and future development trend of intelligent cockpits unilaterally, without fully considering various factors such as drivers, vehicles, and environment. Also, the multi-modal interaction applied in intelligent vehicles has the problem of redundancy, which is not conducive to the driver's operation in terms of convenience and safety. This study focuses on multi-modal interaction, analyzes and evaluates the existing interaction modes of intelligent vehicles. And by combining the driver's cabin experience with the environmental perception of intelligent vehicles, it explores how to enhance the integration of intelligent vehicle cabin and exterior through the creation of a multi-modal interactive fusion AR in car system, forming a closed-loop interaction of "human-vehicle-environment".

2 INTELLIGENT CAR INTERACTION METHODS

2.1 Visual Image Interaction

Any text or material outside the aforementioned margins will not be printed.The application of visual image interaction in intelligent vehicles is very common. The visual interaction equipment inside the cabin consists of input and output terminals. The camera at the input end monitors the driver's posture, face, hands, and other aspects in all directions to detect the driver's emotions or abnormal states (such as signs of fatigue) (Guerrero-Ibáñez et al., 2018). The display screen at the output end provides various feedback, including passenger status, vehicle status, road status, etc, bringing users a nice cabin experience. On the other hand, the visual image interaction outside the cabin reminds the driver of obstacles, pedestrians, traffic lights, and other factors on the road, in order to improve the convenience and safety of driving. In addition, the application of night vision assistance programs (David et al., 2006) outside the cabin has greatly improved the safety level of intelligent vehicles driving at night, helping drivers to more quickly detect potential threats on the road at night. The most important visual image interaction method,

whether inside or outside the cabin, is computer vision (CV). It relies on camera capture and utilizes deep learning algorithm models such as YOLO and CNN to perform real-time analysis and target object detection through the process of image perception, feature extraction, training learning, and classification decision-making (as shown in Figure 1). However, while improving the user experience, visual image interaction still has its immaturity. For example, the feedback provided by the display screen to the driver may cause interference, leading to distraction in real-time road conditions and vehicle driving, and potentially creating safety hazards.

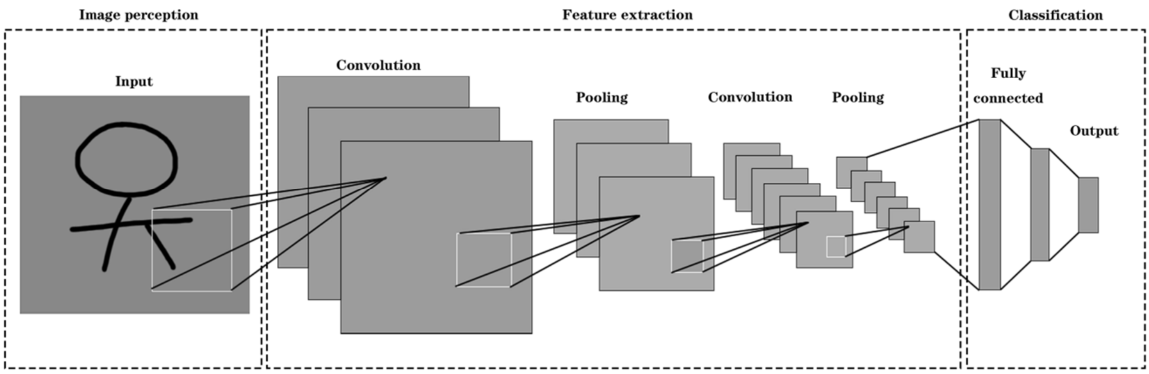

Figure 1: CNN flowchart (Picture credit: Original).

2.2 Touch Interaction

The application of touch interaction in intelligent vehicles is mainly reflected in the cabin, and its main devices are the onboard electronic touch screen and physical touch elements (buttons, knobs, sliders, etc). Vehicle touch screen is an important channel for achieving human vehicle interaction, which can input based on the finger operations of the driver and passengers, and complete feedback and output according to requirements. At the same time, the linkage between the touch screen and other modules (such as audio, Internet of Vehicles (IoV), etc.) enables more convenient functions like car music playback, weather forecasting, map navigation, etc., enhancing practicality and entertainment. The car touch screen has multiple forms. According to technical principles, it includes capacitive touch screens, force feedback touch screens, infrared optical touch screens, etc. They are applied to scenarios with different precision and screen sizes based on different needs. On the other hand, the physical touch elements in vehicles belong to Active Haptics, which require external electrical input to present tactile feedback. Sensors rely on user movements (orthogonal, tangential, or rotational) as shown in Figure 2 to react and trigger actuators (Breitschaft et al., 2019) to control vehicle modules (such as window opening and closing, seat adjustment, and mirror fine-tuning), with high programmability and flexibility. Although touch interaction has greatly improved the convenience, practicality, and entertainment inside the cabin, problems such as accidental touch during bumpy driving, strong lighting affecting the use of the touch screen, and redundant buttons still need to be solved urgently.

Figure 2 Sketch map of haptics (Breitschaft et al., 2019).

2.3 Voice Interaction

In terms of enhancing the cabin experience, visual and touch interactions in smart cars have always had the problem of excessive occupation of the driver's line of sight, which has instead increased the difficulty of operation and reduced driving safety. Therefore, in the field of intelligent vehicles, the applications of natural language interaction based on speech recognition are the trend. Among them, automatic speech recognition (ASR) technology relies on deep learning models, usually built on the architecture of Transformer and RNN, and obtains the voice signals of passengers through steps such as noise reduction enhancement, frame windowing, and feature extraction. The main function of natural language processing (NLP) technology is to perform semantic and sentiment analysis on extracted text to understand user needs and provide feedback. In addition, text to speech (TTS) technology is a branch of artificial intelligence that converts text into smooth and natural speech. Based on end-to-end models such as Transformer and WaveNet, speech synthesis helps drivers focus on the road ahead and reduce the impact of text on the line of sight through speech. The linkage of these several voice technologies constitutes in car voice interaction. It can indeed avoid occupying the driver's line of sight and help them focus on driving. However, research on the interaction interference of in vehicle information systems (IVIS) suggests that these voice based interactions may require high cognitive abilities from drivers, especially older drivers who require a greater workload (Strayer et al., 2016). Meanwhile, multi-modal evaluations of embedded vehicle voice systems indicate that in vehicle voice interaction reduces visual demands, but does not eliminate them (Mehler et al., 2016).

2.4 Other Interaction Modes

Biological signal interaction is a new type of interaction technology for intelligent vehicles. Many traffic accidents are caused by abnormal physical conditions of the driver, so biological signal interaction is particularly important, mainly reflected inside the intelligent cockpit. It non-invasively monitors physiological signals such as electrocardiogram, electroencephalogram, electromyographic response, and respiration, that is, obtains relevant information through indirect means without directly interfering with or disrupting the normal state of the monitored object (Baek et al., 2009). Assessing the health and stress status of the driver, provides prompts and warnings to ensure driving safety. Additionally, the application of Vehicle to Everything (V2X) outside the cabin greatly enhances the interaction between intelligent vehicles and the road environment, including interaction with other vehicles (V2V), interaction with infrastructure such as traffic lights (V2I), interaction with pedestrians (V2P), etc.V2X links smart cars with most factors in the environment, greatly enhancing the car's environmental perception ability.

3 DISCUSSION

3.1 Limitations of Rigid Hybridization in Multi-modal Interaction

The diverse interaction methods mentioned above enhance the car's environmental perception and intelligent cockpit experience, greatly facilitating the driver. However, considering that all functions on smart cars need to meet the premise of "safety", single modal interaction methods have many limitations. In the visual and cognitive evaluation of multi-modal vehicle information systems, researchers analyzed various task types (such as audio entertainment, calling and dialing, text messaging and navigation) and different interaction modes to assess the measurement of user cognitive needs, visual and manual needs, subjective workload, and time spent completing different tasks. It was found that many IVIS functions were too distracting to be activated while the vehicle was in motion (Strayer et al., 2019). Analysis shows that there are also many problems with the rigid combination of current multi-modal interaction methods. Firstly, the complex interaction of various modes has resulted in functional stacking and conflicts, and the priority issues of various interaction modes may lead to operational logic contradictions. Secondly, more interactive methods will also bring higher learning costs, and users need to understand the usage of each modality. At the same time, whether it is visual interaction, touch interaction, or voice interaction, it will distract the driver's attention. How to reasonably grasp the degree of interaction while allowing the driver to concentrate on driving is an urgent problem to be solved. The above issues do not cover everything, but they can basically reflect the limitations of rigid hybridization in multi-modal interaction.

3.2 Key Technologies for Mitigating Problems

Augmented Reality (AR) is a technology that overlays virtual information such as text, images, and 3D models in real time onto the real world, preserving the perception of the real world while displaying virtual information. It significantly reduces the cognitive load of users and extends their senses. Augmented reality is a clear manifestation of "machines adapting to humans", which is in line with the development trend of intelligent cars towards "cars adapting to humans" and "humans adapting to each other".

Based on the series of interaction limitations mentioned above, this article will present a fusion strategy for multi-modal interaction, which complements, balances, and integrates the interaction modes of various modalities from the perspective of augmented reality. This chapter will demonstrate the improvement of vehicle environment perception and cabin experience through the application of several AR technologies in intelligent vehicles using this fusion strategy. Augmented Reality Head Up Display (AR-HUD) directly maps key information in the driving environment, including heading, speed, lane markings, collision warnings, etc., onto the front windshield and presents it in combination with the real road in front of the driver's eyes (Figure 3). Panoramic parking assistance virtualizes the bottom of the vehicle and the road conditions behind it, creating a panoramic image of the blind spot environment, upgrading the vehicle's environmental perception ability and improving parking safety 3D spatial audio, as an auxiliary technology, integrates voice interaction into AR systems, reducing noise interference and improving drivers' speech comprehension ability. Gesture recognition helps users manipulate and grasp virtual information, and compared to touch interaction, it is simpler but more practical. These AR based technologies comprehensively enhance existing visual interaction, touch interaction, voice interaction, and other aspects. They integrate existing interaction methods to form a multi-modal interaction fusion in car AR system, which comprehensively considers the three factors of driver, car cabin, and driving environment, forming a closed-loop interaction of "human car environment" (Figure 4).In this situation, many problems will be easily solved. The display screen (i.e. windshield) in augmented reality seamlessly receives visual information, meeting the driver's visual needs without distracting their attention from the road, thus improving driving safety. Augmented reality enhances the visual interaction between the front, side, and rear of the vehicle, achieving environmental perception visualization, broadening the vehicle's field of view, and filling in perception blind spots. The gesture recognition interaction supported by augmented reality applications reduces the driver's dependence on touch panels and buttons, focuses their operations on the road surface, simplifies operation steps, shortens operation time, and alleviates the complexity of touch interaction. The AR floating window text is directly presented on the front windshield, which greatly reduces the learning cost, alleviates the burden of voice understanding on passengers, and reduces the cognitive pressure of voice interaction.

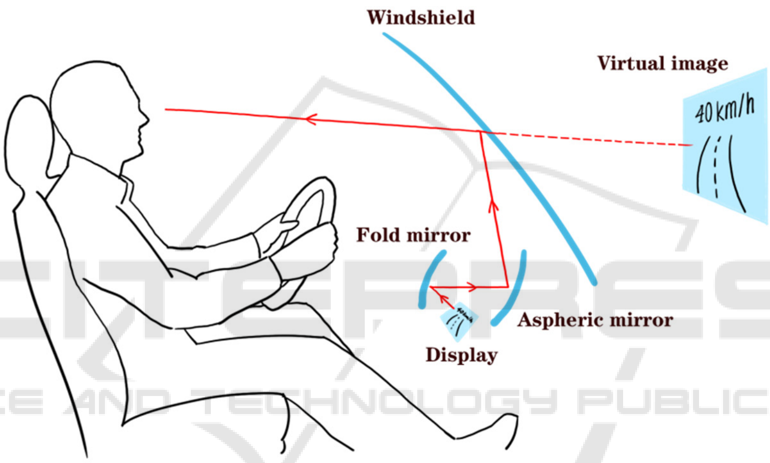

Figure 3: AR-HUD schematic diagram (Picture credit: Original).

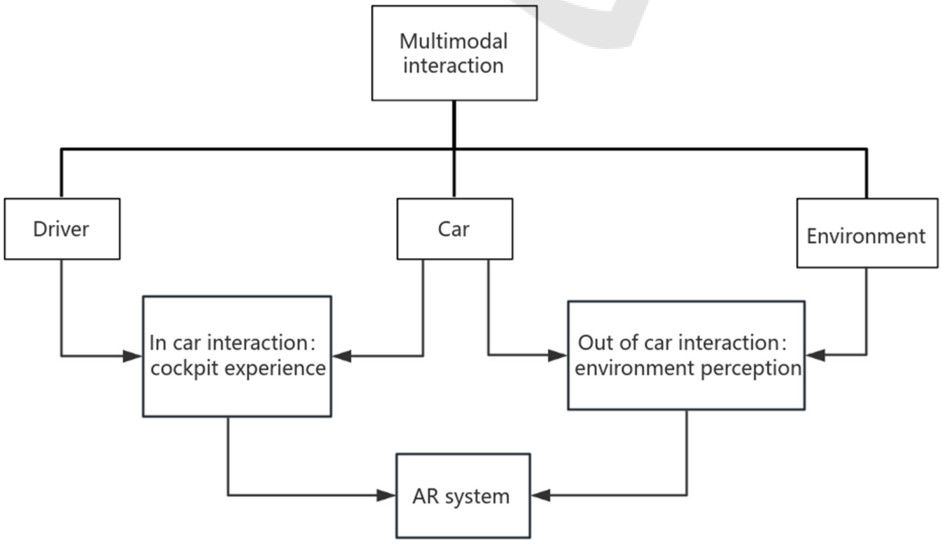

Figure 4: Closed-loop interactive mind map (Picture credit: Original).

3.3 Future Development Challenges

Although the multi-modal interaction fusion in car AR systems has solved many of the above problems, it still has shortcomings. In strong light environments or when driving in reverse, the AR projection is not clear enough, which affects information transmission. In this case, the detection of user gestures may also fail and virtual information cannot be manipulated. Simultaneously, if too much traffic information is projected onto the AR display device or the information layout is not reasonable enough, it will indeed cause information overload problems, which not only hinder the interaction inside the cabin, but also affects the driver's observation of the outside environment. In the future, there are many other AR technologies that can be applied, and other interaction methods can also be integrated into it. There is great room for improvement in this system.

4 CONCLUSIONS

This study discussed how to enhance the environmental perception and cabin experience of intelligent vehicles. From the perspective of multi-modal interaction, the research analyzed various interaction methods such as visual image interaction, touch interaction, and voice interaction to improve the safety, practicality, and comfort of intelligent vehicles. At the same time, the study found the limitations of the single mode interaction mode, as well as the functional conflicts and miscellaneous problems caused by the rigid combination of multi-modal interaction. To address these pain points, this article introduced the application of augmented reality technology in smart cars and concluded that a closed-loop interaction of "human-vehicle-environment" can be achieved by creating an AR in car system that integrates multi-modal interaction, achieving the goal of comprehensively enhancing the perception of intelligent car environment and cabin experience while ensuring safety. These findings highlight the growing importance of integrating emerging technologies to optimize human-vehicle interaction. In the broader context of intelligent transportation systems, future work should continue to explore innovative approaches that bridge technology and user needs, driving the evolution of safer, more intuitive, and more immersive mobility experiences.

REFERENCES

Baek, H. J., Lee, H. B., Kim, J. S., Choi, J. M., Kim, K. K., & Park, K. S. (2009). Nonintrusive biological signal monitoring in a car to evaluate a driver's stress and health state. Telemedicine and e-Health, 15(2), 182-189.

Breitschaft, S. J., Clarke, S., & Carbon, C.-C. (2019). A theoretical framework of haptic processing in automotive user interfaces and its implications on design and engineering. Frontiers in Psychology, 10, 1470.

David, O., Kopeika, N. S., & Weizer, B. (2006). Range gated active night vision system for automobiles. Applied Optics, 45(28), 7248-7254.

Gao, F., Ge, X., Li, J., Fan, Y., Li, Y., & Zhao, R. (2024). Intelligent cockpits for connected vehicles: Taxonomy, architecture, interaction technologies, and future directions. Sensors, 24(16), 5172.

Guerrero-Ibáñez, J., Zeadally, S., & Contreras-Castillo, J. (2018). Sensor technologies for intelligent transportation systems. Sensors, 18(4), 1212.

Mehler, B., Kidd, D., Reimer, B., Reagan, I., Dobres, J., & McCartt, A. (2016). Multi-modal assessment of on-road demand of voice and manual phone calling and voice navigation entry across two embedded vehicle systems. Ergonomics, 59(3), 344-367.

Strayer, D. L., Cooper, J. M., Goethe, R. M., McCarty, M. M., Getty, D. J., & Biondi, F. (2019). Assessing the visual and cognitive demands of in-vehicle information systems. Cognitive Research: Principles and Implications, 4(1), 18.

Strayer, D. L., Cooper, J. M., Turrill, J., Coleman, J. R., & Hopman, R. J. (2016). Talking to your car can drive you to distraction. Cognitive Research: Principles and Implications, 1(1), 16.

Xu, X., Wu, Z., & Zhao, Y. (2024). An improved longitudinal driving car-following system considering the safe time domain strategy. Sensors, 24(16), 5202.

Zhang, Y., Guo, L., You, X., Miao, B., & Li, Y. (2024). Cognitive response of underground car driver observed by brain EEG signals. Sensors, 24(23), 7763.

Zhao, D. (2022). Application of neural network based on visual recognition in color perception analysis of intelligent vehicle HMI interactive interface under user experience. Computational Intelligence and Neuroscience, 2022, 3929110.

Zhou, Y., Fu, R., Wang, C., & Zhang, R. (2020). Modeling car-following behaviors and driving styles with generative adversarial imitation learning. Sensors, 20(18), 5034.