Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 119 (2023) 296–301

www.elsevier.com/locate/procedia

33rd CIRP Design Conference

A Visual Reasoning-based AR-HUD Service Design Approach for Better Driving Experience

Yongshi Liang[1], Pai Zheng[1,2]*, and Liqiao Xia[1]

[1]*Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region, 999077, China*
[2]*Laboratory for Artificial Intelligence in Design, Hong Kong Science Park, New Territories, Hong Kong Special Administrative Region, 999077, China*

* Corresponding author. Tel.: +852 27665633; *E-mail address:* pai.zheng@polyu.edu.hk

Abstract

Smart traffic or transportation, as a critical part of realizing the smart city, utilizes various smart product-service systems to support it. Among them, the augmented reality head-up display (AR-HUD) system is a typical smart driving solution, which projects vital information about driving situations with pre-defined AR graphics onto the windshield of vehicles. The AR-HUD service design attracts increasing concerns owing to its enhancement of drivers' situational awareness without glancing down at the instrument cluster. Nevertheless, current in-car AR-HUDs only stack some simple driving information from sensors, such as the driving speed. It's difficult to meet the demand of drivers since they ignore a higher level of drivers' situation awareness, i.e., driving situation prediction and recommendation of driving advice. To overcome the challenge, this paper proposes a visual reasoning-based service design approach for advancing AR-HUDs' digital servitization and better user experience. Additionally, drivers are involved in the service design development in a value co-creation manner through human-computer interaction design. With the combination of experiential driving knowledge and cognitive intelligence computing on the driving scene, the proposed novel service design approach enables AR-HUD to percept driving scenarios and infer good strategies accordingly. Finally, an illustrative example is carried out to validate the feasibility of the proposed AR-HUD service design approach.

Keywords: Smart product-service system, augmented reality head-up display, visual reasoning, value co-creation, smart driving

1. Introduction

The smart product-service system is widely used to support the construction of smart traffic or transportation, which is a crucial component of smart city. Among them, augmented reality head-up display (AR-HUD) system is a typical smart driving solution, which combines the real driving environment with virtual space. The AR-HUD system can add driving clues to environment or accentuates some elements that already appear in surroundings on the head-up display. Owing to the improvement of drivers' situational awareness (SA) without glancing down at the instrument cluster, the AR-HUD service design is becoming increasingly popular in global auto market, with an estimation of $13.97 billion market sales by 2028.

Moreover, to enhance driver's SA, more recent studies concentrate on the design of AR-HUD service. For example, Bremers et al. [1] explored the perceived difference of perspective depth cues on AR-HUD. For drivers' better understanding of traffic scenes, Rao et al. [2] visualized detected objects with in-car object-level 3D reconstruction. The SA is defined as three levels [3], which are the perception of essential elements in the surroundings (Level 1 SA), events comprehension (Level 2 SA), and projection of future situations (Level 3 SA), respectively. Although there have been some works for the developments of AR-HUD driving service design, most of them focus on displaying some in-car sensor's information (e.g., driving speed) directly and neglect a higher level of driver's SA, namely the driving situation prediction and driving advice recommendation. Therefore,

2212-8271 © 2023 The Authors. Published by Elsevier B.V.

10.1016/j.procir.2023.02.136

these AR-HUDs are difficult to meet the requirement of drivers.

To overcome this challenge, this work proposes a visual reasoning-based service design approach for advancing AR-HUD service with better driving experience. Meanwhile, drivers, as an integral part of the proposed service design architecture, provide their driving experience and knowledge in a value co-creation manner through human-computer interaction (HCI) design. By cognitive intelligence computing, the proposed AR-HUD service design is capable of understanding driving scenarios, while the corresponding driving strategies are recommended based on the experiential driving knowledge of users. The rest of this paper is organized as follows. Section 2 outlines some related works of visual reasoning and AR-HUD. Section 3 describes the proposed AR-HUD service design approach in detail. An illustrative example of the proposed AR-HUD service design is further given in Section 4. At last, Section 5 summarizes the contribution, limitations, and future work.

2. Literature review

2.1. Visual Reasoning

Visual reasoning refers to the ability of a system to carry out multi-hop reasoning over the visual scene's components and their relationship, bridging the gap between scene comprehension and decision-making. The solution of visual reasoning can be divided into the vision-only solution and the language-driven solution. Vision-only solution merely involves visual cues in the reasoning, while language-driven solution additionally introduces language cues. Visual Question Answering (VQA) and Visual Commonsense Reasoning (VCR) are two typical language-driven visual reasoning tasks.

Recently, visual reasoning becoming an attractive approach for cognitive intelligence computing in robotics, computer vision, and other research fields. For example, Murata et al. [4] introduced a long short-term memory (LSTM) based framework to achieve visuomotor reasoning with a given goal image for human-robot collaboration (HRC).

Additionally, according to the study in cognitive computing, reasoning with domain knowledge and human experience can lead to more dependable, efficient, and explicable results. Zheng et al. [5] designed a knowledge memory embedding framework to leverage prior knowledge for visual reasoning. Zheng et al. [6] utilized an HRC knowledge graph (KG) in visual reasoning for a combination of human cognition and robotics in HRC systems. Based on previous research, this paper utilizes a KG consisted visual reasoning approach to design the AR-HUD service solution, improving the digital and intelligent level of in AR-HUD.

2.2. AR-HUD

AR-HUD, as a typical smart driving solution, leverages AR technology to display driving information on the vehicles' windscreen, trying to enhance drivers' SA in the way of combining virtual information and a real environment. More studies have verified that AD-HUD has a beneficial impact on drivers' driving performance [7] and driving cognitive workload [8]. In addition, some explorations in AR-HUD service design have been conducted in previous work. For example, Li et al. [9] proposed an AR-HUD service design to display warming in hazardous situations.

However, the value offered by existing AR-HUDs is limited because their provided information (e.g., lane marking, driving speed, warnings about the lead vehicle, road guidance) is basic and simple. Therefore, this paper attempts to introduce a visual reasoning-based AR-HUD service design approach to deliver higher-level cognitive information (e.g., predicted information and driving strategies) to drivers, thus improving drivers' SA and driving experience. Moreover, this paper considers value co-creation between users and producers in the service development process to gain users' recognition.

3. Methodology

Inspired by the previous work [6], this article proposes a visual reasoning-based service design approach for advancing AR-HUDs' digital servitization and better driving experience. As can be seen in Fig. 1, the proposed AR-HUD service design is outlined with (a) a co-creative value model and (b) a cognitive intelligence computing framework.

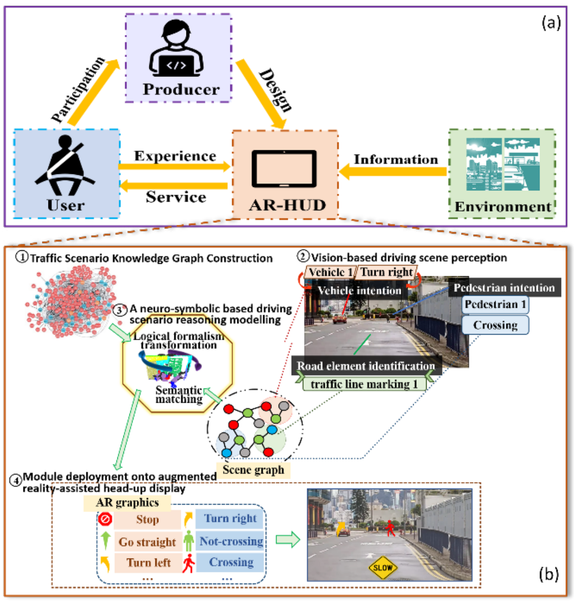

Fig. 1. (a) co-creative value model and (b) cognitive intelligence framework.

Following the service-dominant logic, the proposed service design attempts to co-create value with users. Considering the case in this paper, we pictures a variation of the co-creative value model constructed by Ueda et al. [10] in Fig.1. (a), showing the close interrelation between the user, producer, AR-HUD, and environment.

In addition, the cognitive intelligence computing framework supporting AR-HUD service is shown in (b). Firstly, a KG in the field of driving should be established, which is used to describe various traffic scenarios and store driving knowledge and experience. Next, the scene graph is generated based on vision-based driving scene perception, which includes road elements identification and dynamic elements' intention prediction. And then, a neuro-symbolic driving scenario reasoning model will be constructed by computing the intrinsic relationship between visual perception and pre-defined traffic scenario in KG, giving appropriate driving strategy further. At last, analytical information, including necessary detected road elements in the physical driving environment and appropriate driving strategies, is augmented on the HUD by virtual AR graphics, giving drivers intuitive support. In this way, drivers can enhance their driving SA and reduce cognitive load, achieving better driving experience. More information will be provided below.

3.1. Value Co-creation in AR-HUD Service Design

This paper provides a concrete example of value co-creation in smart product-service systems. The value co-creation mentioned in this article refers to the collaboration with users and producers in the development process of the proposed AR-HUD service design.

Consisting of the user, the producer, the AR-HUD, and the environment, the co-creative model drawn in Fig.1 (a) describes the offered values and the value flow between these containing elements. The producers design the proposed AR-HUD service with the participation of users. As co-developer, users are allowed to take part in the construction of the KG, which is a vital part of the proposed intelligence computing framework, contributing their driving experience to the development of service functionality. According to perceived information from the driving environment, the AR-HUD provides users with corresponding support for smart driving service in the use phase.

Under the value co-creation mechanism, the experiential driving knowledge provided by users can support the completion of the cognitive intelligence computing framework. In the meantime, user participation in the development can clarify the service value, making the proposed AR-HUD service gain more satisfaction from users.

3.2. HCI Design in Driving Scenario Knowledge Graph Construction

A KG that contains the human understanding of the world can drive the computer system toward human-level cognitive intelligence. Meanwhile, design-oriented HCI can contribute to value co-creation in the development of service design since HCI provides an opportunity for user participation, and human cognitive response can come into play when users interact with complex technology. The standard HCI design iteration process includes requirement analysis, alternative design, user study, following up the user's feedback and the problem encountered [11].

In this paper, HCI design is adopted to establish the driving scenario knowledge graph (DSKG), extracting users' driving experience in the development process of AR-HUD service. Specifically, to reveal the driving knowledge and experience of users, users are invited to the driving simulator, taking driving operations according to the driving scenario video given by the computer. With advanced deep learning computing, the driving experience and knowledge of users are extracted in the interaction process and stored as a KG.

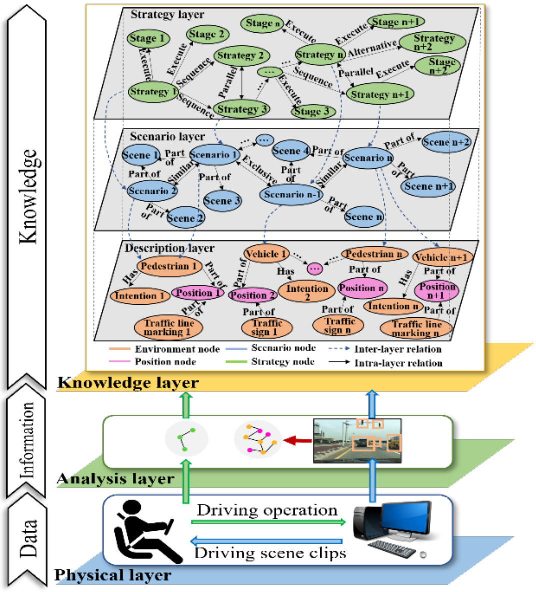

Fig. 2. A technical architecture of HCI design for DSKG construction.

The technical architecture of HCI design, including the physical layer, analysis layer, and knowledge layer, for DSKG construction can be presented in Fig. 2, which shows the user input transfers from data to information, and is stored in DSKG as knowledge.

The physical layer consists of users, a driving simulator, and a computer used for intelligent computing. In the physical layer, according to the driving scene video clips provided by the computer, users take driving operations in the driving simulator based on their driving experience. At the same time, the computer can receive the operation signal from the driving simulator. The operation signals and the corresponding driving scene video are then transmitted to the analysis layer.

In the analysis layer, the computer utilizes the scene graph generation model to perform scene understanding on the driving scenario image captured from the driving scene video. Detailly, A scene graph generation model RelTR [12] with an encoder-decoder architecture can be adopted. The visual features context in the image is coded in the encoder while the decoder output the triples using multiple types of attention networks which are decoupled entity attention, decoupled visual attention, and coupled self-attention. RelTR can generate a scene graph from the driving scenario image directly. Meanwhile, the driving operation signals from users are recognized as the driving strategy entities, linking with the generated scene graph. At last, the information acquired in the analysis layer (the scene graph connecting to the driving strategies entities) is delivered to the knowledge layer and stored in DSKG.

When it comes to DSKG, it is defined as:

$$\mathcal{G} = \{\mathcal{E}, \mathcal{R}, \mathcal{F}\}, \tag{1}$$

where $\mathcal{E}$, $\mathcal{R}$,and $\mathcal{F}$ are entities, relations, and facts respectively. The fact consists of triples:

$$(h, r, t) \in \mathcal{F}, \tag{2}$$

where h and t mean head and tail entities in the group of entities $\mathcal{E}$, and r is a relation in set $\mathcal{R}$.

As shown in the knowledge layer, the DSKG consists of position node, environment node, scenario node, and strategy node which are organized into three layers (description layer, scenario layer, and strategy layer).

The driving experience from users is stored in the strategy layer as different driving operations described by corresponding executive stages. At the same time, the driving scene graph generated from the video clip is represented by dynamic road elements (e.g., vehicles and pedestrians) and static road elements (e.g., zebra crossing, traffic signs, roadblocks, etc.) in the description layer. Moreover, the scenario layer links the strategy layer to the description layer, indicating the users' driving operations under a specific driving scenario. With strategy-oriented reasoning search, the DSKG can provide drivers with a driving suggestion.

In the process of DSKG establishment, HCI design enables the producer to extract users' driving experience and knowledge from their cognition and operative response towards driving scene video. Under the premise of safety and accuracy, driving knowledge and users' experience are incorporated into the cognitive intelligence computing framework in the form of a KG, thus improving performance of the proposed AR-HUD service design and user satisfaction.

3.3. Visual Reasoning-based Driving Scenario Modelling

Visual reasoning-based driving scenario modelling aims to analyze the driving scenario image collected by an in-car camera, identifying the vital road elements and recommending proper driving strategies for drivers. The modelling includes two parts: vision-based driving scene perception and neuro-symbolic driving scenario reasoning, which correspond to ② and ③ in Fig.1 (b), respectively.

The vision-based driving scene perception includes road element identification (e.g., vehicles and pedestrians, traffic signs, traffic lights, etc.) and the behavior intention prediction of dynamic elements, which enables the proposed AR-HUD service to perceive various driving scenarios. Specifically, the proposed service design framework adopts Faster-R-CNN [13], a typical DNN-based object detection model, to identify road elements.

Regarding the behavior intention recognition of dynamic elements, the proposed service design framework considers two main road users, i.e., pedestrians and vehicles. For pedestrians, their behavior intention (walking along in the direction of the vehicle, crossing the road, and not crossing the road) can be recognized by the ST-GCN model [14]. For the intention recognition of nearby vehicles, an encoder-decoder architecture with LSTM networks is used. After feeding the image's features into the model, vehicles' behavior patterns (e.g., stop, turn left, turn right) are predicted.

When it comes to driving scenario reasoning, increasing attention has been paid to the neuro-symbolic model recently, especially in the task of reasoning over KG [15]. As a hybrid model, the neuro-symbolic model equips the data-driven neural network with the abstract logic ability in symbolic processing, leveraging both advantages.

In this paper, based on the neuro-symbolic network, a driving scenario reasoning model is established to select the most appropriate driving strategy in the DSKG under a given driving scenario. The reasoning model consists of two steps: subgraph grounding and logical reasoning. In the subgraph grounding step, the scene graph created from the driving scene perception is transformed into a set of DSKG-aligned candidate triples with semantic matching and a path-based approach. After transforming these DSKG-aligned triples to a formal logic formalism by a rule-based algorithm, a Logical Neural Network (LNN) [16] executes candidate logical queries and inference for giving out an appropriate driving strategy in the logical reasoning step.

Visual reasoning-based driving scenario modelling is a vital part of the cognitive intelligence computing framework in the proposed AR-HUD service design. Visual perception is targeted at recognizing key road elements in the driving scenarios and predicting the behavior intention of the dynamic road elements. The precepted information is then input to the neuro-symbolic driving scenario reasoning model, inferring suitable driving strategies over the DSKG co-established with users.

3.4. Module deployment onto AR-HUD

Being deployed advanced intelligent modules mentioned above, the proposed AR-HUD is able to give out cognitive information (e.g., the predictive behavior intention of pedestrians and vehicles, driving strategy, etc.) to drivers with AR technology. Therefore, the proposed AR-HUD achieves the improvement of drivers' SA and driving experience. Fig. 3 summarizes the transmission of information between the modules in the proposed AR-HUD.

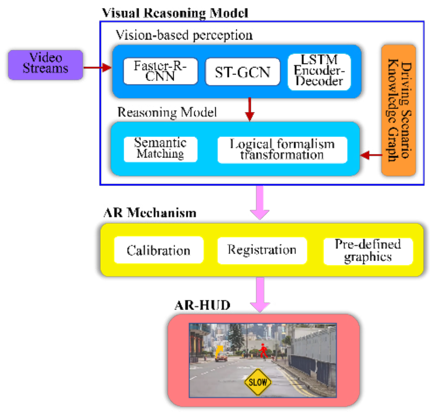

Fig. 3. Module deployment structure of the introduced AR-HUD.

In detail, video streams obtained from the in-car camera are firstly analyzed by the visual perception models, including Fast-R-CNN, LSTM-based encoder-decoder model, and ST-GCN. The perceptual outcome as the form of a scene graph is then fed into the reasoning model, which infers appropriate driving strategies over the DSKG by semantic matching and logical formalism transformation. After that, the perceptual outcome and the recommended driving suggestion are transmitted to the AR mechanism. Finally, through calibration and registration in the AR mechanism, these analyzed results can be suitably rendered onto the head-up display with pre-defined graphics.

4. An illustrative example

The AR-HUDs in the current market cannot satisfy the demand of drivers to some extent since they only stack some basic information (e.g., driving speed) on display. In a value co-creation manner with users and producers, the proposed AR-HUD service is designed to provide drivers with more cognitive information, such as road elements identification, dynamic elements' behavior estimation, and timely recommended driving strategies, aiming to decrease drivers' cognitive workload and increase driving experience. An illustrative example that gradually achieves driving scenarios perception and strategies reasoning is given to demonstrate the viability of the proposed AR-HUD service design.

4.1. The configuration of hardware and software

As shown in Fig. 4, the proposed AR-HUD service design is deployed in the simulated indoor driving environment.

The hardware consists of a driving simulator with Logitech G29 Driving Force, a Wi-Fi connected LCD projector playing driving video stream on the screen for showing driving road condition, a 55-inch see-through OLED display as a windshield, and a computer with an Intel (R) Core (TM) i5 processor and 16 GB RAM for data analysis and AR realization. Especially, unlike the existing in-car AR-HUD using the general HUD gadget, this paper utilizes a larger see-through OLED display as a HUD device in the driving simulation experiment, expecting to widen the driver's field of view and deliver more useful information.

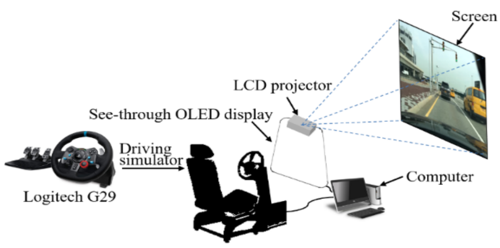

Fig. 4. Hardware configuration of the simulated indoor experiment.

Regarding software, the software development kit (SDK) released by Logitech is applied to read the underlying signals from users' operation in the driving simulator for realizing HCI design. Besides, this paper uses Neo4j to visualize the DSKG constructed in this example. Moreover, the cognitive intelligence computing framework in the proposed AR-HUD service design is established with Visual Studio 2019 installed OpenGL and OpenCV library in the Windows 10 operating system. The utilization of OpenCV is to conduct driving scenario perception, and the AR realization relies on OpenGL.

4.2. Stepwise implementation

The implementation of the proposed AR-HUD service design basically involves four steps, which are aligned with the principal procedures indicated in Fig. 1(b).

Step 1: DSKG Establishment through HCI design. From the perspective of value co-creation, the DSKG is firstly constructed in collaboration with users through HCI design. Some details are given below:

1) Participants: to ensure the quality of DSKG, users taking part in the co-development process should have a driving license with rich driving experience.

2) Apparatus: as shown in Fig.4, the driving simulator maintains the fidelity of the real driving cockpit as much as possible. In the experiment, the driving scenario video is played on the screen with the projector, and the data is collected with the Logitech SDK.

3) Driving scenario videos: according to the normal reaction speed of a driver on the road, the driving scenario video is about 5 seconds in length. In addition, the videos used in the experiment cover various driving scenarios as much as possible to improve the granularity of the KG.

In the experiment, after an introduction about the experiment, participants are invited to sit in the driving simulator alone and asked to take driving operations based on their driving experience and knowledge when the video is playing. The videos and the driving operations of participants (i.e., driving strategies which are turn right, turn left, move forward, and slow down) are then read by computer and analyzed with algorithms such as scene graph generation model RelTR, thus completing the construction of DSKG. At last, cross validation and sample inspection are taken into consideration for the dependability of DSKG, guaranteeing driving strategy associated with each driving scenario is clear and appropriate.

Step 2: Holistic driving scenario perception. The detection of road components and the behavior prediction of dynamic road components are both included in driving scenario perception. Take the typical road condition in Fig. 1 (b) as an example. The road components (a traffic line marking, a motor vehicle, and a pedestrian) and their location are detected. Additionally, a 9-layer ST-GCN model based on transfer learning predicts the pedestrian will cross the street, while a 4-layer LSTM-based encoder-decoder model forecasts the car will turn right. The perceptual outcomes are then organized in the scene graph for the subsequent process.

Step 3: Neuro-symbolic driving scenario reasoning. The target of driving scenario reasoning is to seek the most trustworthy driving strategy from a group of driving strategy nodes in the DSKG under a specific driving scenario. The scene graph generated in Step 2 stores the holistic perceptual results. For subgraph grounding, the scene graph is firstly aligned with the DSKG via semantic matching, including entity linking and relation linking (similarity threshold of 0.7). Next, a path-based algorithm is used to retrieve a set of DSKG-aligned candidate triples by exploring the paths in the scene graph. After that, these DSKG-aligned candidate triples are converted to a standard first-order logic representation via a rule-based algorithm. The transformed logic representation enables the LNN model to conduct logical inference, which recommends slowing down in the given road condition.

Step 4: Showcasing analyzed information on AR-HUDs. After driving scenarios perception in Step 2 and driving strategies reasoning in Step 3, the analyzed information, such as detected road components, predicted behavior intention, and inferred driving strategies, is showcased on the HUD with AR graphics, thereby providing more intuitive support and better driving experience for drivers.

4.3. Discussion.

The visual reasoning-based AR-HUD service design approach can be implemented stepwise by following the steps outlined above, which realizes the delivery of cognitive information to drivers with AR technology. Noteworthily, the proposed service design approach can equip with vehicle-to-everything technology further, enabling the vehicle to share analyzed information with the vehicle nearby. For instance, if the view of drivers, especially the truck driver, is blocked, the blind spot message based on the proposed AR-HUD's analytical information from surrounding motor vehicles can be transmitted to the drivers. Therefore, as an artificial intelligence-based driving aid for better driving experience, the introduced AR-HUD service design is evolvable, high-tech integrated, and promising.

5. Conclusion and future work

Although the AR-HUD system has been adopted as a promising smart driving solution, the service provided by most existing AR-HUDs is hard to meet drivers' satisfaction with a low level of SA. To fill the gap, this paper proposes a visual reasoning-based service design approach for delivering perceptive, prophetic, and inferential information on the AR-HUD at the same time, allowing drivers' higher-level SA and better driving experience. To our knowledge, this is the first attempt to equip cognitive intelligence technology with AR-HUD. To verify the feasibility of the proposed approach, an illustrative example under a specific driving scenario was conducted, and the main contribution of this work can be summarized as follow:

1) A domain-specific DSKG was constructed through HCI design, which realizes value co-creation between users and producers in the service design development process and increases user recognition of the provided service.

2) A novel AR-HUD service design approach based on visual reasoning was firstly proposed for driving situation prediction and driving advice recommendation, enhancing driving experience.

Nevertheless, some limitations still exist. The cognitive intelligence computing framework used in the proposed AR-HUD service design consists multiple modules stacked successively. For example, the prediction of behavior intention relies on the result of road element perception, while the reasoning of driving scenario is based on the road elements detection and behavior intention prediction. It means the error will accumulate between modules and long-time consumption is inevitable. Meanwhile, the DSKG designed in this study is not comprehensive enough to include all road elements, leading to the disability of complicated traffic scenario reasoning.

Thereby, future work should be further carried out to ensure the proposed AR-HUD service design success: 1) an end-to-end model should be developed for the enhancement of robustness and timeliness, and 2) the DSKG should be further enriched and better organized with multi-granularity and various traffic scenario characteristics.

Acknowledgements

The work described in this paper was supported by the Smart Traffic Fund of the Transport Department (Ref. PSRI/35/2202/PR), HKSAR, China.

References

[1] Bremers, A.W.D., Yöntem, A.Ö., Li, K., Chu, D., Meijering, V., Janssen, C.P., 2021. Perception of perspective in augmented reality head-up displays. International Journal of Human-Computer Studies 155, 102693.

[2] Rao, Q., Chakraborty, S., 2021. In-Vehicle Object-Level 3D Reconstruction of Traffic Scenes. IEEE Transactions on Intelligent Transportation Systems 22, 7747–7759.

[3] Endsley, M.R., 2017. Toward a theory of situation awareness in dynamic systems. Human Error in Aviation 37, 217–249.

[4] Murata, S., Masuda, W., Chen, J., Arie, H., Ogata, T., Sugano, S., 2019. Achieving human--robot collaboration with dynamic goal inference by gradient descent. In: International Conference on Neural Information Processing. pp. 579–590.

[5] Zheng, W., Yan, L., Gou, C., Wang, F.Y., 2021. KM4: Visual reasoning via Knowledge Embedding Memory Model with Mutual Modulation. Information Fusion 67, 14–28

[6] Zheng, P., Li, S., Xia, L., Wang, L., Nassehi, A., 2022a. A visual reasoning-based approach for mutual-cognitive human-robot collaboration. CIRP Annals 71, 377–380.

[7] Medenica, Z., Kun, A.L., Paek, T., Palinko, O., 2011. Augmented Reality vs. Street Views: A Driving Simulator Study Comparing Two Emerging Navigation Aids. In: Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services, MobileHCI '11. Association for Computing Machinery, New York, NY, USA, pp. 265–274.

[8] Janner, M., Narasimhan, K., Barzilay, R., 2018. Representation learning for grounded spatial reasoning. Transactions of the Association for Computational Linguistics 6, 49–61.

[9] Li, Z., Ma, Y., 2021. Response and evaluation of vehicle AR-HUD assistant system to risk cognition. ICCSE 2021 - IEEE 16th International Conference on Computer Science and Education 153–157.

[10] Ueda, K., Takenaka, T., Váncza, J., Monostori, L., 2009. Value creation and decision-making in sustainable society. CIRP annals 58, 681–700.

[11] Mohammed, Y.B., Karagozlu, D., 2021. A Review of Human-Computer Interaction Design Approaches towards Information Systems Development. BRAIN Broad Research in Artificial Intelligence and Neuroscience 12, 229–250.

[12] Cong, Y., Yang, M.Y., Rosenhahn, B., 2022. RelTR: Relation Transformer for Scene Graph Generation 1–14.

[13] Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems 28.

[14] Yan, S., Xiong, Y., Lin, D., 2018. Spatial temporal graph convolutional networks for skeleton-based action recognition. In: Thirty-Second AAAI Conference on Artificial Intelligence.

[15] Amizadeh, S., Polozov, A., Palangi, H., Huang, Y., Koishida, K., 2020. Neuro-Symbolic Visual Reasoning: Disentangling "Visual" from "Reasoning." Pmlr 1–28.

[16] Riegel, R., Gray, A., Luus, F., Khan, N., Makondo, N., Akhalwaya, I.Y., Qian, H., Fagin, R., Barahona, F., Sharma, U., Others, 2020. Logical neural networks. arXiv preprint arXiv:200613155.