Study on Human-Computer Interaction System of Vehicles

車両におけるヒューマン・コンピュータ・インタラクション・システムに関する研究

Wenping Luo

Saitama Institute of Technology

February, 2022

Contents

Abstract

This paper deals with the control of vehicles and wheelchairs using a human-computer interaction (HCI) system based on artificial intelligence technology. The proposed system learns and recognizes biological signals (eye gaze, blink, brain waves, etc.) and converts the corresponding commands to the computer into various communication protocols such as serial, USB, TCP/IP, etc. The system can perform actions according to the corresponding commands. A combination of signal processing and image recognition and other techniques are used to extract the biological signals. Similarly, image recognition is also used for some of the obstacle detection decisions. In order to accurately determine dynamic obstacles in real time, our system uses single shot detection method. The proposed HCI system was experimentally validated by applying it to an electric wheelchair and an electric vehicle.

When the system was applied to a wheelchair for experimental purposes, eye tracking and blink detection were used to achieve the wheelchair motion. By extracting the pupil of the eye image by binarization and localizing the pupil centre, the system captures the trajectory of eye movement and determines the direction of gaze. On the other hand, we build a convolutional neural network for feature extraction and classification of open and closed eye images, and perform machine learning to extract features from multiple independent images of open and closed eye states and input them into the system. The system also relies on the P300 neurophysiology protocol, which uses EEG signals to drive the wheelchair. While driving, the subject confronts the stimulator, a real-time virtual orientation interface, and pays attention to the commands to be executed. This visual stimulation triggers neural phenomena and a target area is detected by EEG signal processing. This target area represents the direction to drive the wheelchair to the desired location. By using this as a brain-computer interface, the HCI system is more than 90% accurate. In this study, we have validated the HCI system on five healthy subjects. Based on these results, a technical evaluation and variability study of the device is reported, showing good adaptability with all users being able to use the device successfully with relative ease.

Furthermore, the proposed system could be used for autonomous driving and safety assistance in future vehicles. In order to improve the reliability of the system, we implanted it into a small Toyota electric car for experimental validation. A convolutional neural network was built and used for single shot detection and blink detection. The single shot detection method was used to achieve dynamic target detection. For blink detection, multiple images of eyes open and closed were input to the network for deep learning, and based on the learned data model, the open/closed state of the subject's eyes could be detected. Eye tracking technology is also used to identify the driver's gaze direction. The proposed system can accurately detect external dynamic targets, and the driver can control the vehicle by blinking or gazing in the direction.

Keywords: human-computer interaction, binarization, convolutional neural networks, machine learning, single shot detection, blink detection, eye gaze, wheelchair, EV.

Chapter 1. Introduction

The trend of automotive intelligence and networked connectivity continues to deepen the digital transformation of the automotive industry, bringing new impacts on the relationship between people, vehicles and the environment, and human-computer interaction design has become a core element of intelligent vehicle development and innovation [1-9]. A series of functions around intelligent autonomous driving system, in-vehicle infotainment system, intelligent in-vehicle voice interaction system and road condition information system will establish a special ecosystem with the vehicle as a carrier to realize the scenario-based application and expansion of intelligent vehicles. To this end, this report develops the research on the current situation and development trend of human-computer interaction of intelligent vehicles, and provides reference for future intelligent vehicle human-computer interaction design [5-9].

In the driving environment, most of the external information is obtained through vision, and other parts are from sound and body perception. The safety of autonomous driving could be greatly enhanced if human biological signals could be applied to the field of autonomous driving [10-15]. It is a powerful aid for people with severely limited motor skills. Of these, eye-movement recognition is again the most worthwhile to delve into, as it allows them to operate the computer using only eye and eyelid movements. Thus, eye movement recognition devices are useful in improving the quality of life for people with severe disabilities [16-20]. With the continuous development of science and technology, artificial intelligence technology will follow the gaze driving technology in various human-computer interaction systems and control systems, such as autonomous driving, robotics, language recognition, image recognition, natural language processing, and expert systems, to be the main protagonists of this autonomous driving technology [21-25]. For this reason we can integrate the characteristics of these technologies and develop a multi-modal recognition HCI system to solve these problems. As for human-computer interaction, it is the study of communication and communication between human and computer through mutual understanding to accomplish the functions of information management, service and processing for people to the maximum extent, so that the computer can really become a harmonious assistant for people to work and study.

1.1 Development of human-computer interaction system

The starting point of human-computer interaction is the text user interface (TUI) interaction. Computers first existed only in the laboratory, the mainframe was huge and sold at a very high price, and those who used computers were professionally trained experts. The text interface is the original appearance of the computer system, it consists of thousands of command lines. These command lines are the first step in the human-computer dialogue, the user through the input command line to complete the communication with the computer [26-29]. However, it has obvious disadvantages, first of all, the interaction steps are long and tedious, you need to enter complex commands to communicate with the computer, the user in the process is very easy to produce a sense of fatigue; second, the learning cost is high, in addition to a few computer experts after professional learning, most of the ordinary people simply cannot smoothly enter and read these command lines. In order to reduce the learning cost, simplify the interaction process, so that most people can also smoothly use the computer, human-computer interaction ushered in its first revolution - the graphical user interface (GUI) interaction [29-33]. With the rapid development of artificial intelligence technology, human-computer interaction technology and artificial intelligence technology began to combine in depth, and human-computer interaction from the original graphical interface interaction gradually to voice interaction, image recognition, eye recognition, gesture recognition, brain-computer interface and other directions, the use of these interaction methods to make people's life more convenient. Therefore, more humanized and intelligent human-computer interaction has become a new theme in the development of human-computer interaction.

1.1.1 Graphical User Interface

Graphical user interface is the user interface of computer operation displayed graphically, which is the dialogue interface between the computer and its users and is an important part of the computer system. In a graphical user interface, windows, icons, buttons and other graphics are displayed on the computer screen to indicate actions for different purposes, which are selected by the user with a pointer device such as a mouse. [29-33]

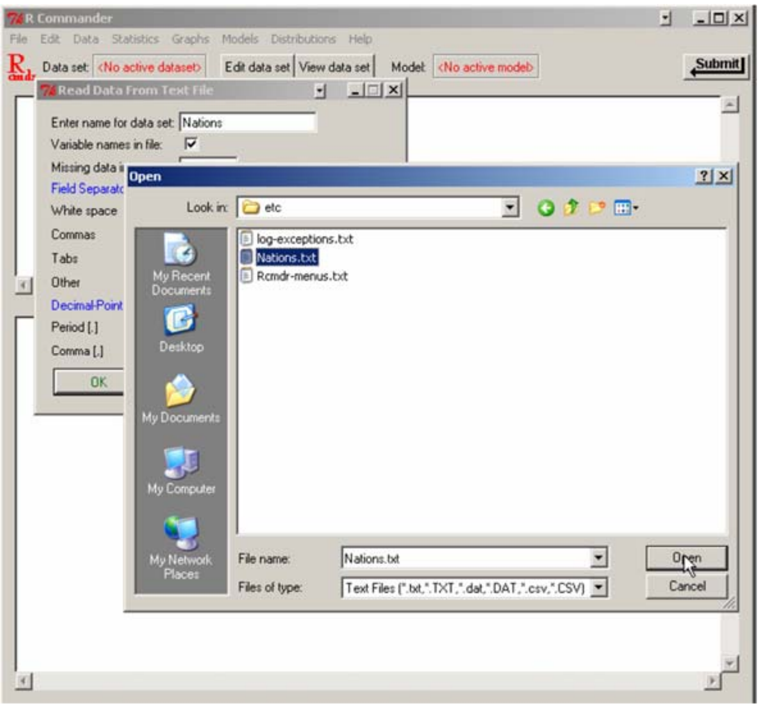


Figure 1-1. Graphical user interface: Reading data from a text file [30]

1.1.2 Gesture recognition interaction technology

Gesture recognition is a new type of natural human-computer interaction technology that integrates advanced perception technology and computer pattern recognition technology. By recognizing human gestures, computers allow users to use simple gestures to interact with devices without directly touching them [34-38]. The original gesture interaction was to use devices positioned on the hand and elbow to detect the trajectory of hand movement and thus achieve the effect of interaction. This method of relying on external devices for interaction, although highly accurate, largely hindered the natural expression of people's gesture activities, and then computer vision-based gesture recognition and interaction technologies were born.

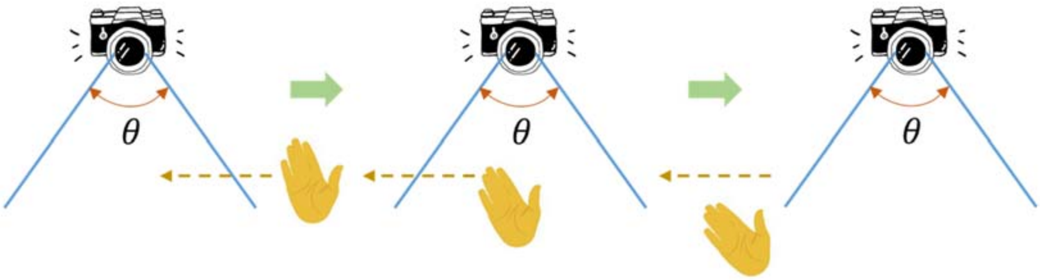


Figure 1-2. Computer vision-based gesture recognition and interaction technology

1.1.3 Voice recognition interaction technology

The earliest voice interaction is the interactive voice response (Interaction Voice Response) system, with which the user interacts by dialing a telephone number. However, this kind of interactive voice response system has a narrow application range, low interaction efficiency and rigid interaction mode [39-43]. Due to the many drawbacks of interactive voice response system and its inability to solve many practical problems of users, mobile APPs such as Siri, Google, Xiao Ai classmates, which integrate visual and voice interaction, and pure voice interaction intelligent products such as Amazon Echo and Xiao Du were born. The birth and success of these products not only proved the value of voice interaction, but also accelerated the development of voice interaction technology.

Figure 1-3. Voice recognition and interaction technology

1.1.4 Brain Computer Interface technology

Brain Computer Interface (BCI) is a multidisciplinary human-computer interface involving neuroscience, cognitive science, computer science, control and information science technology, and medicine. It is a neural information exchange and control channel established between the brain and the external environment, allowing the brain to exchange signals in both directions between the computer and external devices [8] [44-48]. BCI technology can accurately and quickly collect and identify brain signals from the human brain under various thinking activities, and use these signals to control external devices.

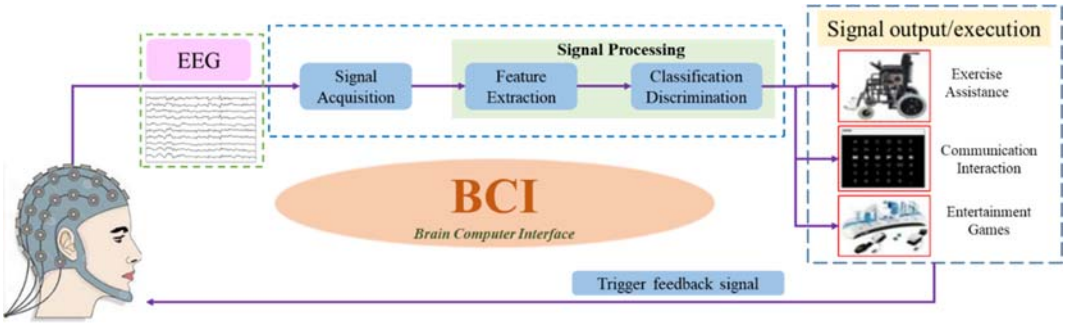


Figure 1-4. Schematic diagram of the principle of brain-computer interface

1.2 The HCI system based on eye movement recognition.

Eye movement detection technology is the detection of eye and eye movements to determine where the user is looking. Eye movement starts by finding the part of the eye that is not moving (reference point) and the part that is moving (moving point). Once the reference point and the moving point are found, gaze is detected based on the position of the moving point relative to the reference point. In general, there are two eye-tracking techniques: the first method measures the position of the head compared to the eye, and the second measures the point of interest of the eye in space [49-53]. HCI systems focus on the user's object of interest in the interaction scene and usually use post-use measurements. The most widely used measurement point is the gaze tracking method based on the pupil-corneal reflection vector. While eyelid movements we can use machine learning and depth methods for feature classification of open-eye and closed-eye images so that blinks can be detected. In this study, A driving control system using eye gaze and blink input devices has been developed for people with severe disabilities.

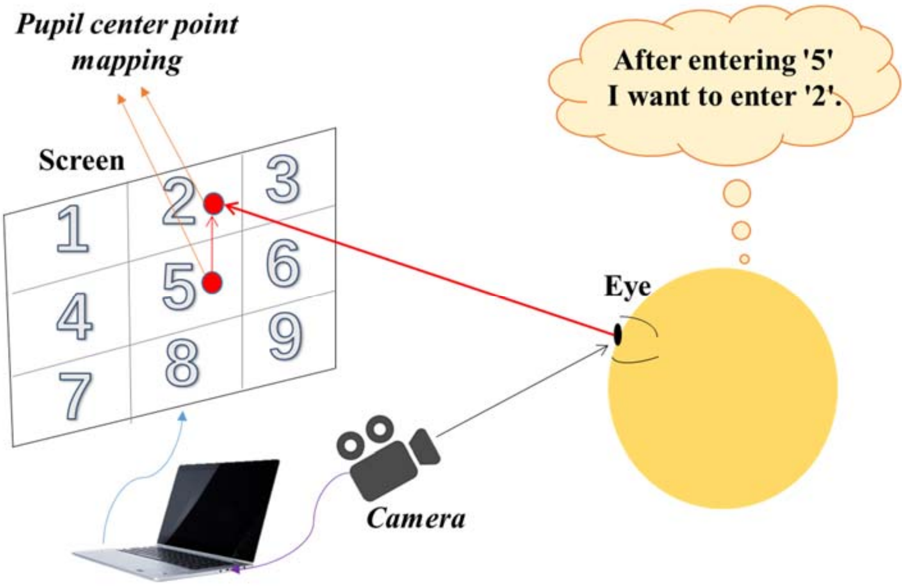


Figure 1-5. Principle of eye tracking technology

1.3 The HCI in vehicle system

Human-computer interaction is a comprehensive discipline, and the interaction design of intelligent vehicles involves a number of research fields such as vehicle engineering, ergonomics, psychology, and computer science. The research of vehicle engineering includes the study of the relationship between vehicle power system, vehicle braking system, vehicle navigation system, vehicle multimedia system and driver as well as the human-vehicle-road model; the research of industrial design includes the design of display mode (graphics, color, display device), the design of operation device (button shape, layout, operation device structure, shape) in human-computer interaction system; the research of psychology includes the study of human-computer interface design based on Ergonomics research includes the relationship between human, machine and environment in the "human-machine-environment" system, and provides theories and methods to solve the problems of human effectiveness and health in the system; computer science research includes the study of the development of software and hardware for human-computer interaction systems. Computer science research includes the development of software and hardware for human-computer interaction systems, image recognition and processing, natural language processing, data analysis and decision making [1-4].

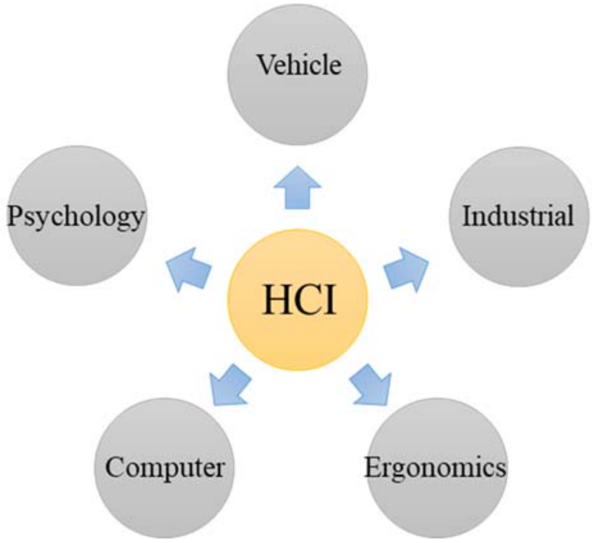


Figure 1-6. Human-computer interaction involves research areas

The deepening intelligence of automobiles makes the operation of each function more complicated and brings new challenges to human-computer interaction design. The human-computer interaction design of intelligent vehicles must be based on ensuring the safety of the vehicle, improving the efficiency of the operation of core functions and simplifying the operation process, so as to improve the user experience of the vehicle's occupants and give full play to the functions and performance of the vehicle [55-59].

The design process of HCI system mainly includes demand analysis stage, investigation and research stage, system analysis and planning stage, system design stage, testing stage, production and manufacturing of HCI system and submission for use stage. In the design process, the characteristics and needs of the system's service recipients are clarified through research and analysis, and the product planning and design work is carried out with the user in mind, including the interaction mode, the shape and size of each component, hardware and software, etc., which is put into use after the testing of the relevant items is completed.

The future smart driving car will realize customized human-computer interaction design based on usage scenarios. The interior design of smart driving cars will be redefined and developed in the direction of customization and multi-functionalization because of the simplification of the manipulation mechanism and the large amount of space that can be released. The increase of functions makes the human-computer interaction system of smart driving cars have more room for development. The human-computer interaction design concept of traditional cars with driving as the core element will be weakened, and the realization of multiple scenarios will become the main consideration of human-computer interaction design issues. Compared with traditional cars, the in-vehicle information system of smart driving cars will also be more powerful, supporting the car to achieve a variety of functions such as autonomous driving or assisted driving, status monitoring, entertainment, office and communication.

Multi-functional and multi-channel interaction design makes smart driving cars tend to develop in the direction of customization. The interaction design concept, priority and design style of autonomous driving function, multimedia entertainment function, mobile office function and intelligent navigation function in smart driving cars will be more dependent on the definition of scenarios, and the interaction mode will also tend to be multi-channel due to the expansion of space and function, realizing the integration of intelligent voice interaction, gesture control, touch screen interaction and physical button interaction. The scenario will drive the smart car and interaction design to specific function combinations based on usage orientation, for example, the smart car equipped with mobile office system and mobile network can become a mobile office and mobile meeting room; the smart car equipped with multimedia entertainment system and mobile network can become a mobile rest and entertainment room [1-9] [55-60].

At present, most automotive companies have launched human-machine interaction systems, and the interaction methods are basically based on touch screen, physical buttons/knobs, voice control and other methods. Based on the different design concepts of each manufacturer, the design of control methods, operation processes and control areas have their own characteristics, and some manufacturers support in-depth services such as background manual remote control and after-sales support.

1.3.1 HCI applications for vehicles

Human-computer interaction based on touch, hearing, vision and other multi-modalities is the future development trend of intelligent vehicles. With the maturity of various intelligent technologies, the integration of various forms of interaction such as visual interaction, voice interaction and gesture interaction can be realized, and the cooperative interaction of different functions can be realized according to different scenes. Smell and taste interaction has not become the mainstream interaction mode at present, nor has it been applied in mass production vehicles on a large scale. Multimodal HCI-based systems are used in several aspects of vehicles, such as: autonomous driving systems, wearable devices, elderly welfare devices, small electric vehicles, etc., as shown in Figures 1-6.

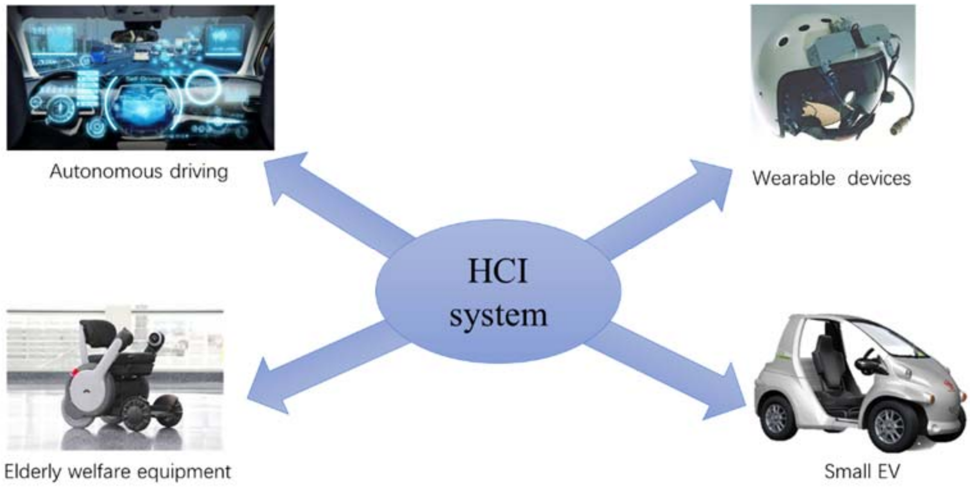


Figure 1-7. Multimodal human-computer interaction in vehicles

1.3.2 Vehicle HCI system to obtain environmental information

Intelligent vehicle human-computer interaction will be designed with "human-vehicle-environment" as a whole, while scenario-driven in-vehicle design towards customization. For the era of intelligent networked vehicles, the vehicle is only a part of the entire transportation system, intelligent vehicle interaction design needs to take into account the top-level design of the entire transportation system, not only to improve the efficiency of the interaction between the driver and passenger in the vehicle as the goal, but also to consider the optimal integration of the transportation system and the environment.

Future human-computer interaction is not only a single interaction between people and vehicles, but also a cross-transport transportation system composed of people, vehicles, infrastructure, cities and the environment. Vehicle intelligence will be more reflected in the V2X aspect, vehicle autopilot and environmental intelligence interoperability for the vehicle to bring a new functional expansion, not only to the vehicle itself intelligent driving, real-time environmental data analysis, intelligent navigation, energy replenishment strategy and other functional enhancements, but also the intelligent upgrade of the entire urban transportation system, which helps to alleviate urban traffic congestion, urban vehicle monitoring and management, and Driving behavior and violation information statistics, etc.

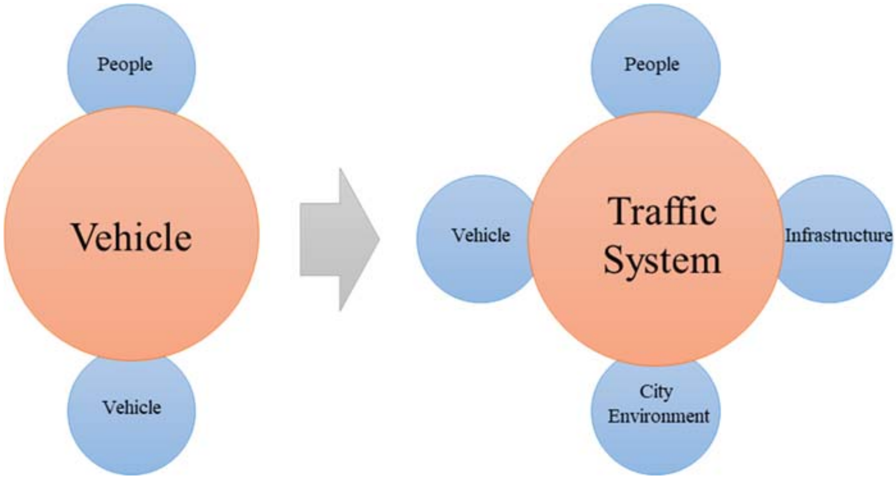


Figure 1-8. Trends in the interaction between vehicles and the environment

1.3.3 Emotional interaction based on intelligent technology

Intelligent emotional interaction is an important trend and ultimate embodiment of human-computer interaction in automobiles. With the development of artificial intelligence technology and biometric technology, intelligent cars will have human-like perception, thinking and behavior functions, with the ability of emotion recognition, emotion understanding and emotion expression. At the same time, the scope of interaction will also be expanded, able to interact with the passengers inside the car and with the pedestrians outside the car, more intelligent and humane. For example, when the driver inside the car is fatigued, the car can alert the driver through vibration, sound or light, and guide to the nearest parking lot to rest or start the automatic driving mode. When the car occupants are recognized to be seasick, the driver will be reminded to adjust the driving style or turn on the interior ventilation and other functions to realize the emotional execution of the function.

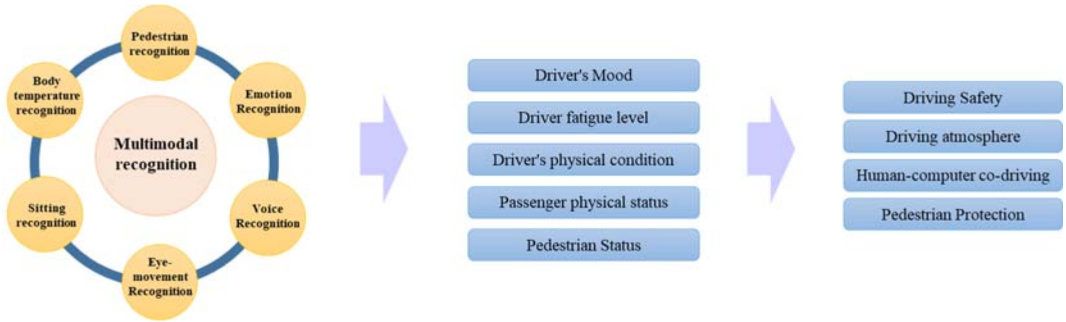


Figure 1-9. Interaction of multimodal recognition and information in vehicles

1.3.4 HCI in vehicles with the autonomous driving system

The vehicle's autonomous driving technology includes video cameras, radar sensors, and laser rangefinders to understand the surrounding traffic and navigate the road ahead with an exhaustive map. All of this is accomplished through a data center that processes the vast amount of information gathered by the vehicle about the surrounding terrain. In this respect, self-driving vehicles are the equivalent of remote-controlled cars or smart cars in data centers [61-76]. However, the maps and road information in the data center are collected in advance by a human driver, and if the data for a destination is not collected, that destination cannot be reached by an autonomous vehicle. At this time, At this point, we need to control the vehicle to reach the destination through the vehicle's human-machine interaction system, drive the vehicle through manual combined with the vehicle's assisted driving system, while collecting relevant data and uploading it to the data center.

Figure 1-10. The vehicle with automatic driving system [54]

The future smart car will serve as a platform to integrate a large number of functions, and human-computer interaction plays a key role in the safety and operational efficiency of the car. Any car is operated or used by people, and the efficiency, safety, convenience and even the psychological pleasure and fatigue of people in operating the equipment designed by the human-computer interaction system directly affect the safety and working condition of the car. It is necessary to give full play to the respective characteristics and advantages of human and machine as much as possible, adopt scientific and effective interaction methods, make maximum use of the machine, and at the same time give full play to the active role of people and Wisdom, so that man and machine work together to form a reasonable human-computer interaction system. In addition, with the improvement of the level of intelligent driving and commercialization of cars, the trend of scenario-based cars becomes more prominent, and the human-machine interaction design based on different scenarios will meet more functional needs of users. The new human-computer interaction will change the existing driving mode of the car. Smart driving cars have different modes of artificial driving, assisted driving and unmanned driving, including the conversion between human and vehicle, the realization of task takeover and handover, task transfer and processing of emergencies, and information interaction of smart driving vehicles and integrated information management platforms. Human-vehicle cooperative driving will become the core interaction design direction of intelligent driving vehicles, and the interaction mode, interaction process, interaction safety and interaction efficiency will become the key research objects.

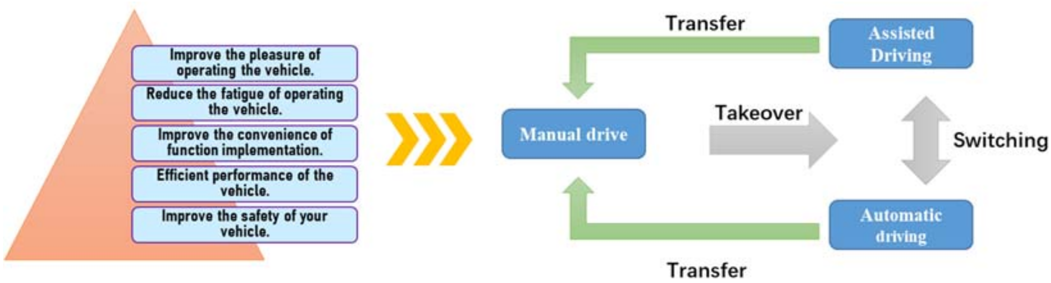


Figure 1-11. Multimodal human-computer interaction for driving systems

In order to make vehicle HCI systems more intelligent, HCI systems with multimodal recognition technologies have been applied to automobiles, of which eye recognition and speech recognition are mostly the most popular. Therefore, in this study, a multimodal HCI system based on eye recognition was developed and applied to a wheelchair for experimental validation. In addition, in order to ride the wave of technological innovation, automakers around the world are focusing on the development of autonomous driving by conducting long-distance and extended test drives on public roads. If development progresses well, self-driving capabilities are expected to be available within a few years. In the process, there are concerns about the safety of self-driving cars, as it is important to recognize traffic signs and regulate driving because self-driving cars must follow the same traffic rules as drivers. Therefore, it is necessary to detect traffic signs quickly and accurately during the autonomous driving process. For this reason, this study uses single detection as a method for detecting traffic signs. Feature extraction and classification are performed using convolutional neural networks (CNN) in deep learning. In this process, the image is first read and a feature map is created by convolution or ensemble, and these features are used to detect the target. In this study, a novel multichannel human-computer interaction system was developed by combining eye movement recognition and obstacle detection, and it was experimentally validated by applying it to COMS EV.

1.4 The purpose and contribution

1.4.1 Developed a multimodal HCI system for wheelchairs

A wheelchair with multimodal human-computer interaction system has been developed to facilitate mobility of ALS patients and people with hand and foot disabilities. Eye gaze detection and blink detection are included in the system, and the subject can also control the wheelchair by eye movement.

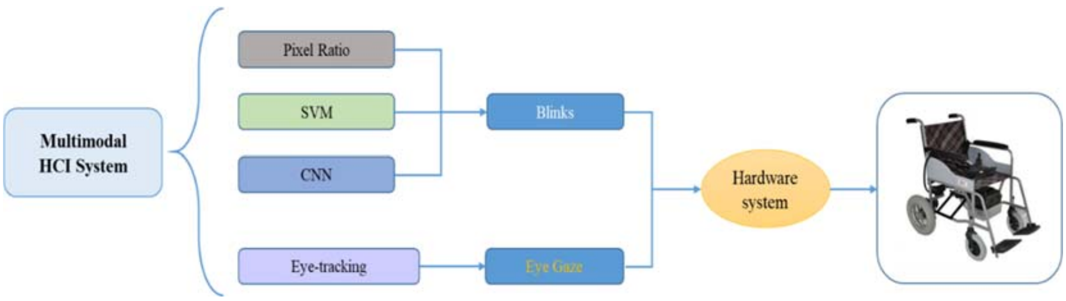


Figure 1-12. Multimodal Human-Computer Interaction System for Wheelchairs

1.4.2 Developed a multimodal HCI system for EV

Combine the multimodal human-computer interaction with autonomous driving and use it as an assisted driving safety system. The system includes eye gaze detection, blink detection and obstacle detection, and it can drive automatically in some good working conditions, and the subject can control the vehicle by eye movement.

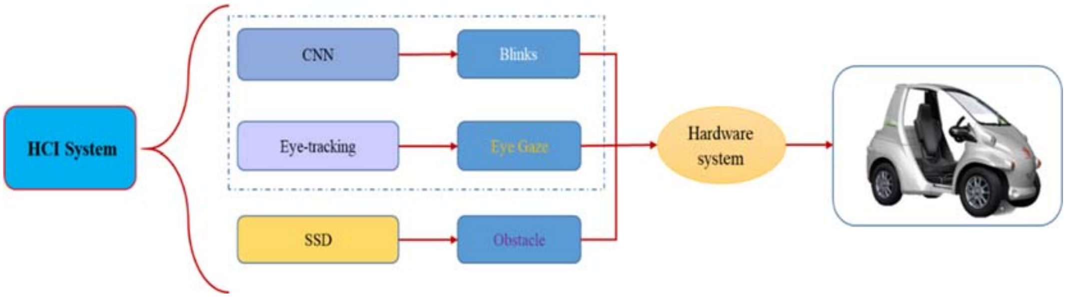


Figure 1-13. Multimodal Human-Computer Interaction System of EV

1.4.3 Blink detection

Blink detection can detect drowsiness and attention of the driver, or switch gears by blinking. Subjects could communicate the appropriate information by the number and timing of blinks.

Table 1. Three blink recognition methods

Methods	Characteristics
Pixel ratio	The fastest detection, requiring only one comparison image and no complex calculation and learning process.
SVM	Solves the problem of small-sample machine learning in high-dimensional space.
CNN	With a certain amount of sample set, the accuracy of recognition can be greatly improved after learning.

1.4.4 Eye gaze detection

The basic working principle of eye tracking is to use image processing technology, using a special camera that can lock the eye to continuously record changes in vision, track the frequency of visual gaze and the duration of gaze, and according to this information to analyze the tracked person. The eye tracking camera, which can be placed in front of a computer screen or set on the screen, can record the user's visual gaze shift with the help of infrared technology and sample recognition software. Eye gaze input is a system that allows the user to enter characters on the screen by gazing at them, and it was developed primarily as a communication aid for people with severe physical disabilities. According to the numeric keypad, the image of the eyeball's moveable range is divided into 5 areas, which are labeled 2, 4, 5, 6 and 8. The 5 areas will match the relative commands.

1.4.5 Designed hardware system for wheelchair

In order to be able to drive the wheelchair through an interactive system, we designed and developed a PIC-based driver circuit for the wheelchair, on which we used a wireless communication serial port so that wireless communication could be made between the computer and the wheelchair.

1.4.6 Raspberry Pi based embedded system in EV

We designed a hardware circuit system for EV with Raspberry Pi as the core and some servo motor drive circuits, and used TCP/IP to communicate with the developed HCI system.

1.4.7 Obstacle detection

Traffic signs and traffic signals can be recognized and fed back to the vehicle while autonomous driving. When using eye gaze to control the vehicle, it can identify obstacles and provide emergency warnings to the driver. We are able to detect the distance of obstacles by our proposed geometric algorithm.

References

[1] Schmidt, Albrecht, et al. "Automotive user interfaces: human computer interaction

in the car." CHI'10 Extended Abstracts on Human Factors in Computing Systems (2010): 3177-3180.

[2] Wang, Kang. "Human-Computer Interaction Design of Intelligent Vehicle-Mounted Products Based on the Internet of Things." Mobile Information Systems, 2021.

[3] Gao, Han, Bo Li, and Ting Han. "User Experience Research on the Human-Computer Interaction System of Connected Car." International Conference on Applied Human Factors and Ergonomics. Springer, Cham, 2019.

[4] Cavett, Doran, et al. "Human-computer interface for control of unmanned aerial vehicles." 2007 IEEE Systems and Information Engineering Design Symposium. IEEE, 2007.

[5] Shaikh, Asadullah. "An Interactive Design using Human Computer Interaction for Autonomous Vehicles." International Journal of Engineering Trends and Technology (IJETT) 10.1 (2020): 160-172.

[6] Galarza, Eddie E., et al. "Real time driver drowsiness detection based on driver's face image behavior using a system of human computer interaction implemented in a smartphone." International Conference on Information Technology & Systems. Springer, Cham, 2018.

[7] Krömker, Heidi, and Cindy Mayas. "Introduction to the Special Issue on Human-Computer Interaction in Autonomous Vehicle and Mobility." International Journal of Human–Computer Interaction 37.4 (2021): 295-296.

[8] Tiwari, Ravikumar K., and S. D. Giripunje. "Design approach for EEG-based human computer interaction driver monitoring system." Int J Latest Trends Eng Technol IJLTET 3.4 (2014): 250-255.

[9] Das, Subasish, and Hamsa Zubaidi. "Autonomous Vehicles and Pedestrians: A Case Study of Human Computer Interaction." International Conference on Human-Computer Interaction. Springer, Cham, 2021.

[10] Kim, Kyu-Beom, et al. "Study on the Variation of Driver's Biosignals According to the Color Temperature of Vehicle Interior Mood Lighting." Science of Emotion and Sensibility 23.2 (2020): 3-12.

[11] Nacpil, Edric John Cruz, Zheng Wang, and Kimihiko Nakano. "Application of

physiological sensors for personalization in semi-autonomous driving: A review." IEEE Sensors Journal, 2021.

[12] Lee, Heung-Gu, Dong-Hyun Kang, and Deok-Hwan Kim. "Human–Machine Interaction in Driving Assistant Systems for Semi-Autonomous Driving Vehicles." Electronics 10.19 (2021): 2405.

[13] Bojarski, Mariusz, et al. "Visualbackprop: Efficient visualization of cnns for autonomous driving." 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018.

[14] Kim, Ki-Hong, et al. "A Biosignal-Based Human Interface Controlling a Power-Wheelchair for People with Motor Disabilities." ETRI journal 28.1 (2006): 111-114.

[15] Prabakaran, Bharath Srinivas, et al. "BioNetExplorer: Architecture-Space Exploration of Biosignal Processing Deep Neural Networks for Wearables." IEEE Internet of Things Journal 8.17 (2021): 13251-13265.

[16] LaCourse, John R., and Francis C. Hludik. "An eye movement communication-control system for the disabled." IEEE Transactions on Biomedical Engineering 37.12 (1990): 1215-1220.

[17] Zheng, Xiaoxiang, et al. "A portable wireless eye movement-controlled human-computer interface for the disabled." 2009 ICME International Conference on Complex Medical Engineering. IEEE, 2009.

[18] Utaminingrum, Fitri, et al. "Eye movement as navigator for disabled person." Proceedings of the 2016 International Conference on Communication and Information Systems, 2016.

[19] Gautam, Gunda, et al. "Eye movement based electronic wheel chair for physically challenged persons." Int. J. Sci. Technol. Res 3.2 (2014).

[20] Mathew, Sharon, et al. "Eye Movement Based Cursor Control and Home Automation for Disabled People." 2019 International Conference on Communication and Electronics Systems (ICCES). IEEE, 2019.

[21] Muhammad, Khan, et al. "Deep learning for safe autonomous driving: Current challenges and future directions." IEEE Transactions on Intelligent Transportation Systems 22.7 (2020): 4316-4336.

[22] Sallab, Ahmad EL, et al. "Deep reinforcement learning framework for autonomous driving." Electronic Imaging 2017.19 (2017): 70-76.

[23] Ning, Huansheng, et al. "A Survey on Hybrid Human-Artificial Intelligence for Autonomous Driving." IEEE Transactions on Intelligent Transportation Systems, 2021.

[24] Grigorescu, Sorin, et al. "A survey of deep learning techniques for autonomous driving." Journal of Field Robotics 37.3 (2020): 362-386.

[25] Shafaei, Sina, et al. "Uncertainty in machine learning: A safety perspective on autonomous driving." International Conference on Computer Safety, Reliability, and Security. Springer, Cham, 2018.

[26] Goodisman, Aaron Aaron Abraham. A Stylus-Based User Interface for Text: Entry and Editing. Diss. Massachusetts Institute of Technology, 1991.

[27] Shneiderman, Ben, Donald Byrd, and W. Bruce Croft. "Sorting out searching: A user-interface framework for text searches." Communications of the ACM 41.4 (1998): 95-98.

[28] Drori, Offer. "The user interface in text retrieval systems." ACM SIGCHI Bulletin 30.3 (1998): 26-29.

[29] Jansen, B. J. (1998). The graphical user interface. *ACM SIGCHI Bulletin*, *30*(2), 22-26.

[30] Fox, J. (2005). Getting started with the R commander: a basic-statistics graphical user interface to R. J Stat Softw, 14(9), 1-42.

[31] Naressi, A., Couturier, C., Devos, J. M., Janssen, M., Mangeat, C., De Beer, R., & Graveron-Demilly, D. (2001). Java-based graphical user interface for the MRUI quantitation package. Magnetic resonance materials in physics, biology and medicine, 12(2-3), 141.

[32] Binosi, D., & Theussl, L. (2004). JaxoDraw: A graphical user interface for drawing Feynman diagrams. Computer Physics Communications, 161(1-2), 76-86.

[33] Oulasvirta, A., Dayama, N. R., Shiripour, M., John, M., & Karrenbauer, A. (2020). Combinatorial optimization of graphical user interface designs. Proceedings of the IEEE, 108(3), 434-464.

[34] Sánchez-Nielsen, Elena, Luis Antón-Canalís, and Mario Hernández-Tejera. "Hand gesture recognition for human-machine interaction." (2004).

[35] RAUTARAY, Siddharth S.; AGRAWAL, Anupam. Vision based hand gesture recognition for human computer interaction: a survey. Artificial intelligence review, 2015, 43.1: 1-54.

[36] TURK, Matthew; ATHITSOS, Vassilis. Gesture recognition. Computer vision: A reference guide, 2020, 1-6.

[37] SUAREZ, Jesus; MURPHY, Robin R. Hand gesture recognition with depth images: A review. In: 2012 IEEE RO-MAN: the 21st IEEE international symposium on robot and human interactive communication. IEEE, 2012. p. 411-417.

[38] MITRA, Sushmita; ACHARYA, Tinku. Gesture recognition: A survey. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 2007, 37.3: 311-324.

[39] Kisilevsky, B. S., Hains, S. M., Lee, K., Xie, X., Huang, H., Ye, H. H., ... & Wang, Z. (2003). Effects of experience on fetal voice recognition. Psychological science, 14(3), 220-224.

[40] Kisilevsky, B. S., Hains, S. M., Lee, K., Xie, X., Huang, H., Ye, H. H., ... & Wang, Z. (2003). Effects of experience on fetal voice recognition. Psychological science, 14(3), 220-224.

[41] SHELDON, Aron, et al. Putting the AR in (AR) chitecture-Integrating voice recognition and gesture control for Augmented Reality interaction to enhance design practice. 2019.

[42] BIRCH, B.; GRIFFITHS, C. A.; MORGAN, A. Environmental effects on reliability and accuracy of MFCC based voice recognition for industrial human-robot-interaction. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 2021, 09544054211014492.

[43] OMYONGA, Kevin; SHIBWABO, Bernard Kasamani. The application of real-time voice recognition to control critical mobile device operations. 2015.

[44] AUNGSAKUN, Siriwadee, et al. Robust eye movement recognition using EOG signal for human-computer interface. In: International Conference on Software

Engineering and Computer Systems. Springer, Berlin, Heidelberg, 2011. p. 714-723.

[45] BULLING, Andreas, et al. Eye movement analysis for activity recognition using electrooculography. IEEE transactions on pattern analysis and machine intelligence, 2010, 33.4: 741-753.

[46] BULLING, Andreas; ROGGEN, Daniel. Recognition of visual memory recall processes using eye movement analysis. In: Proceedings of the 13th international conference on Ubiquitous computing. 2011. p. 455-464.

[47] SHI, Da, et al. Identification and Influence of Tourism Rituals: Analysis of Eye Movement Recognition of Tourism Images. Journal of Hospitality & Tourism Research, 2021, 1096348021996783.

[48] DONG, Wentao, et al. ANFIS fusion algorithm for eye movement recognition via soft multi-functional electronic skin. Information Fusion, 2021, 71: 99-108.

[49] VALLABHANENI, Anirudh; WANG, Tao; HE, Bin. Brain—computer interface. In: Neural engineering. Springer, Boston, MA, 2005. p. 85-121.

[50] WANG, Yijun, et al. A practical VEP-based brain-computer interface. IEEE Transactions on neural systems and rehabilitation engineering, 2006, 14.2: 234-240.

[51] VAUGHAN, Theresa M., et al. Brain-computer interface technology: a review of the Second International Meeting. IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society, 2003, 11.2: 94-109.

[52] VAN GERVEN, Marcel, et al. The brain–computer interface cycle. Journal of neural engineering, 2009, 6.4: 041001.

[53] SILVERSMITH, Daniel B., et al. Plug-and-play control of a brain–computer interface through neural map stabilization. Nature Biotechnology, 2021, 39.3: 326-335.

[54] https://www.cnbeta.com/articles/tech/1081645.htm

[55] Jianan, Lyu, and Ashardi Abas. "Development of Human-Computer Interactive Interface for Intelligent Automotive." International Journal of Artificial Intelligence 7.2 (2020): 13-21.

[56] Xu, Wei. "User Centered Design (VI): Human Factors Approaches for Intelligent Human-Computer Interaction." arXiv preprint arXiv:2111.04880 (2021).

[57] DanNuo, Jiao, et al. "Design of Intelligent Vehicle Multimedia Human-Computer Interaction System." IOP Conference Series: Materials Science and Engineering. Vol. 563. No. 5. IOP Publishing, 2019.

[58] Várkonyi-Kóczy, Annamária R., and Balázs Tusor. "Human–computer interaction for smart environment applications using fuzzy hand posture and gesture models." IEEE Transactions on Instrumentation and Measurement 60.5 (2011): 1505-1514.

[59] Liu, Yujia, et al. "Human-Computer Collaborative Interaction Design of Intelligent Vehicle—A Case Study of HMI of Adaptive Cruise Control." International Conference on Human-Computer Interaction. Springer, Cham, 2021.

[60] Kashef, Mohamad, Anna Visvizi, and Orlando Troisi. "Smart city as a smart service system: Human-computer interaction and smart city surveillance systems." Computers in Human Behavior 124 (2021): 106923.

[61] Levinson, Jesse, et al. "Towards fully autonomous driving: Systems and algorithms." 2011 IEEE intelligent vehicles symposium (IV). IEEE, 2011.

[62] Huang, Xinyu, et al. "The apolloscape dataset for autonomous driving." Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2018.

[63] Caesar, Holger, et al. "nuscenes: A multimodal dataset for autonomous driving." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.

[64] Chen, Chenyi, et al. "Deepdriving: Learning affordance for direct perception in autonomous driving." Proceedings of the IEEE international conference on computer vision. 2015.

[65] Urmson, Chris, et al. "Autonomous driving in urban environments: Boss and the urban challenge." Journal of field Robotics 25.8 (2008): 425-466.

[66] Sallab, Ahmad EL, et al. "Deep reinforcement learning framework for autonomous driving." Electronic Imaging 2017.19 (2017): 70-76.

[67] Bender, Philipp, Julius Ziegler, and Christoph Stiller. "Lanelets: Efficient map representation for autonomous driving." 2014 IEEE Intelligent Vehicles Symposium Proceedings. IEEE, 2014.

[68] Kong, Jason, et al. "Kinematic and dynamic vehicle models for autonomous driving

control design." 2015 IEEE intelligent vehicles symposium (IV). IEEE, 2015.

[69] Chen, Xiaozhi, et al. "Multi-view 3d object detection network for autonomous driving." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2017.

[70] Yoneda, Keisuke, et al. "Automated driving recognition technologies for adverse weather conditions." IATSS research 43.4 (2019): 253-262.

[71] Chen, Siheng, et al. "3d point cloud processing and learning for autonomous driving: Impacting map creation, localization, and perception." IEEE Signal Processing Magazine 38.1 (2020): 68-86.

[72] Cui, Yaodong, et al. "Deep learning for image and point cloud fusion in autonomous driving: A review." IEEE Transactions on Intelligent Transportation Systems (2021).

[73] Yabuuchi, Kento, et al. "Visual Localization for Autonomous Driving using Pre-built Point Cloud Maps." 2021 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2021.

[74] Cho, Sungjin, et al. "Semantic Point Cloud Mapping of LiDAR Based on Probabilistic Uncertainty Modeling for Autonomous Driving." Sensors 20.20 (2020): 5900.

[75] Wu, Pengxiang, Siheng Chen, and Dimitris N. Metaxas. "Motionnet: Joint perception and motion prediction for autonomous driving based on bird's eye view maps." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.

[76] Yang, Kaihong, Sheng Bi, and Min Dong. "Lightningnet: Fast and accurate semantic segmentation for autonomous driving based on 3D LIDAR point cloud." 2020 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2020.

[77] Cadena, Cesar, et al. "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age." IEEE Transactions on robotics 32.6 (2016): 1309-1332.

[78] Kaess, Michael. "Simultaneous localization and mapping with infinite planes." 2015 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2015.

[79] Saeedi, Sajad, et al. "Multiple-robot simultaneous localization and mapping: A review." Journal of Field Robotics 33.1 (2016): 3-46.

[80] Tovar, Benjamín, et al. "Planning exploration strategies for simultaneous localization

and mapping." Robotics and Autonomous Systems 54.4 (2006): 314-331.

[81] Kaess, Michael. "Simultaneous localization and mapping with infinite planes." 2015 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2015.

[82] Jain, Saksham, et al. "Simultaneous Localization and Mapping for Autonomous Robot Navigation." 2021 International Conference on Communication, Control and Information Sciences (ICCISc). Vol. 1. IEEE, 2021.

[83] Silveira, Geraldo, Ezio Malis, and Patrick Rives. "An efficient direct approach to visual SLAM." IEEE transactions on robotics 24.5 (2008): 969-979.

[84] Leutenegger, Stefan, et al. "Keyframe-based visual–inertial odometry using nonlinear optimization." The International Journal of Robotics Research 34.3 (2015): 314-334.

[85] Bloesch, Michael, et al. "Iterated extended Kalman filter based visual-inertial odometry using direct photometric feedback." The International Journal of Robotics Research 36.10 (2017): 1053-1072.

[86] Tanskanen, Petri, et al. "Semi-direct EKF-based monocular visual-inertial odometry." 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2015.

[87] Sun, Ke, et al. "Robust stereo visual inertial odometry for fast autonomous flight." IEEE Robotics and Automation Letters 3.2 (2018): 965-972.

[88] Tong, Wang, et al. "Artificial intelligence for vehicle-to-everything: A survey." IEEE Access 7 (2019): 10823-10843.

[89] Wang, Jian, et al. "A survey of vehicle to everything (V2X) testing." Sensors 19.2 (2019): 334.

[90] Hasan, Monowar, et al. "Securing vehicle-to-everything (V2X) communication platforms." IEEE Transactions on Intelligent Vehicles 5.4 (2020): 693-713.

Chapter 2. Methods

In this chapter, the aim is to provide a description of general methods, mainly some image preprocessing methods and classification methods used in the study. In addition, this includes details of some methods that could be used in future studies and may be considered necessary if this work is to be reproduced or continued in future studies.

2.1 Image Binarization Processing

Binarization of an image is the process of graying the points on the image to 0 or 255, which means that the whole image is rendered in a distinct black and white effect. The grayscale image with 256 brightness levels is selected by appropriate thresholding to obtain a binarized image that can still reflect the overall and local characteristics of the image. In digital image processing, binary image occupies a very important position, especially in the practical image processing, the system composed of binary image processing is a lot, to carry out the processing and analysis of binary image, first of all, the grayscale image should be binarized to obtain the binarized image, which is beneficial when doing further processing of the image, the nature of the image set is only related to the pixel value of 0 or 255 points of This makes the processing easier and the amount of data processing and compression smaller. To obtain the ideal binary image, closed, connected boundaries are generally used to define regions that do not overlap. All pixels with a grayscale greater than or equal to the threshold are determined to belong to a specific object, indicated by a grayscale value of 255, otherwise these pixel points are excluded from the object region, and a grayscale value of 0 indicates a background or exceptional object region. Figure 2-1 shows a grayscale image. The threshold is set and then binarised to give a picture with only black and white.

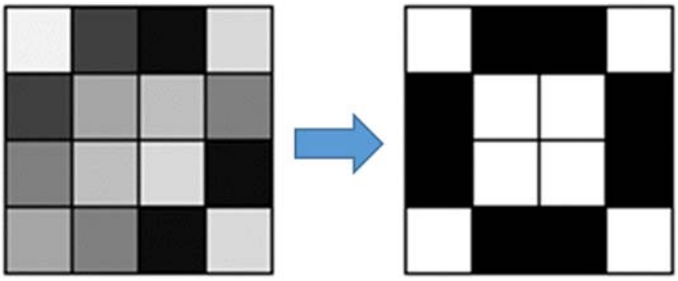


Figure 2-1. Binarization of the grayscale image

If a particular object has a uniform and consistent gray value inside and it is in a uniform background with other levels of gray values, a comparable segmentation effect can be obtained using the threshold method. If the difference between the object and the background is not expressed in grayscale values (e.g., different textures), this difference feature can be converted to a difference in grayscale and then the image can be segmented using threshold selection techniques. Dynamic adjustment of the threshold value to achieve the binarization of the image can dynamically observe the specific results of its segmented image.

$$f'(x, y) = \begin{cases} 255 & f(x,y) \geq \theta \\ 0 & f(x,y) < \theta \end{cases} \tag{2-1}$$

Where the threshold is θ, the coordinate system is (x, y), the concentration value is $f(x, y)$, and the transformed concentration value is $f'(x, y)$.

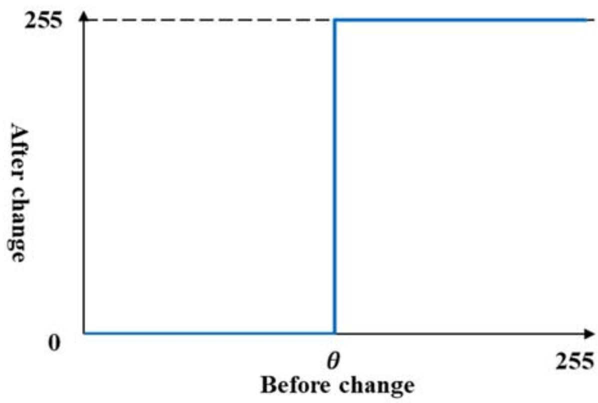


Figure 2-2. Binarization curve with set threshold

Figure 2-3 shows that first the threshold value is set and then after binarization, the color image on the left becomes a black and white image, but the contours are still clear and highly recognizable, which reduces the dimensionality of the image and reduces the computational effort, but retains the main amount of features of the image.

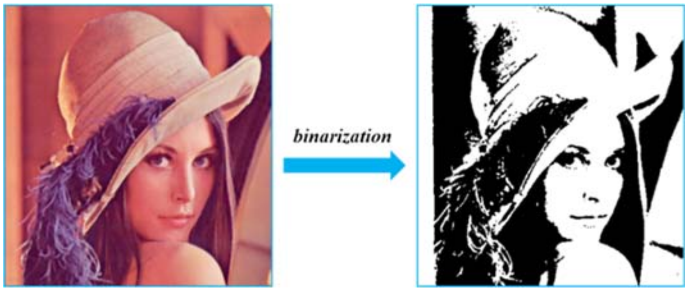


Figure 2-3. Binarization of the color image

2.2 Pixel Ratio

A pixel is a small square of an image that is made up of small squares that have a definite location and a color value assigned to them, and the small square color and location determine how that image appears. A pixel can be thought of as an indivisible unit or element of an entire image. Indivisible means that it cannot be cut into smaller units or elements, but exists as a single color cell. Each dot matrix image contains a certain number of pixels, which determine the size of the image on the screen. Since the pixel values are 0 to 255, the image pixel ratio is the ratio of the value of each pixel in the image multiplied by the number of pixels, to 255 multiplied by the number of pixels.

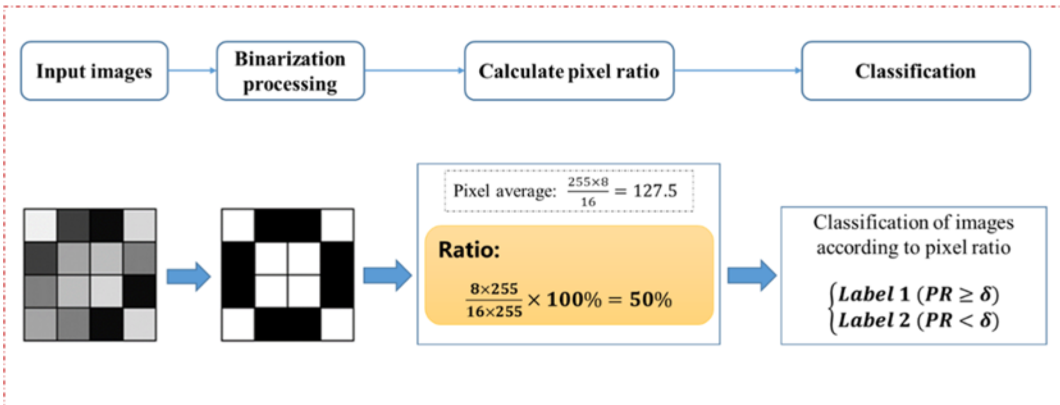


Figure 2-4. Pixel ratio image classification process

We introduce the variance value of the image concentration to compare the difference of the images, the larger the variance value, the more obvious the image difference. First, the average value of each pixel in each row of the image is calculated in turn, and second, the variance of all the obtained averages is calculated, and the obtained variance is the feature value of the image. Finally, the difference of variance values between two images is then compared.

$$\sigma^2 = \frac{\sum(X-\mu)^2}{N}$$

(2-2)

Here, σ^2 is the overall variance, X is the sample concentration value, μ is the sample average concentration value, and N is the overall number of samples.

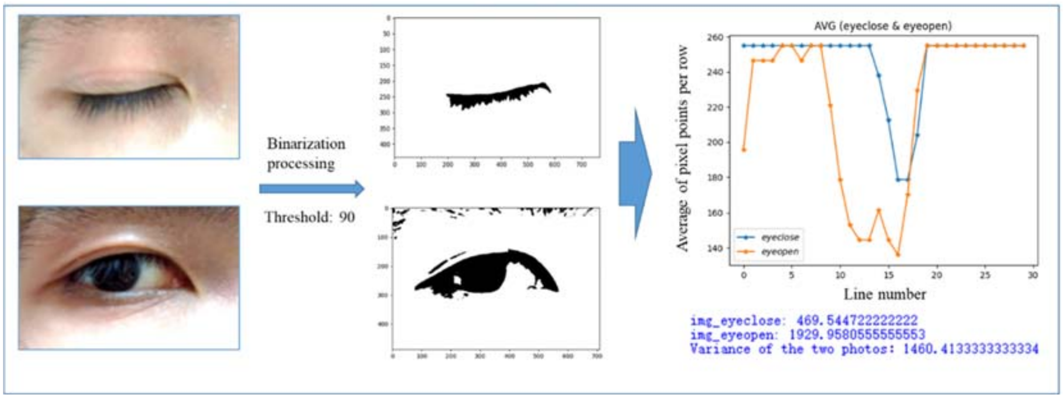


Figure 2-5. Comparison of variance values without background images

2.3 Support vector machine

The support vector machine is finding the best separated hyperplane (Decision Hyperplane above). The best separated hyperplane in the current space is not ideal, and can be found by mapping the data from the original space to a higher dimensional space through the mapping function, but this will increase the amount of operations, in order to reduce the amount of operations, the "Kernel Trick" is introduced; the ideal state of finding the hyperplane is that the two types of data can be completely separated, that is, there is no In practice, "soft spacing" is introduced to achieve a hyperplane that allows for mis-segregated training samples, considering "more practical" separation or training samples that are not completely separable in the space itself; finally, the hyperplane can be solved directly using quadratic To reduce the computational effort, the problem is transformed using lagrangian duals and finally solved quickly using the SMO algorithm.

As in Figure 2-6, there are two types of points in the sample, blue circles and yellow triangles, and there are many lines to separate these two types of points, but the best separation is only one, achieving the maximum interval Margin. The best separation *wx+b=0* in the figure is a hyperplane (*w* is a vector). The support vector machine is to solve for *w* and *b* that form that hyperplane, where the vector represented by the nearest point to the hyperplane *wx+b=0* is called the support vector, i.e. the vector represented by all the points on the two hyperplanes |*wx+b*|=*1*.

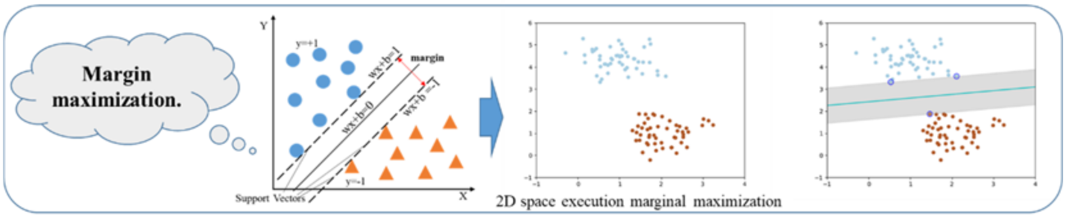


Figure 2-6. 2D space execution marginal maximization

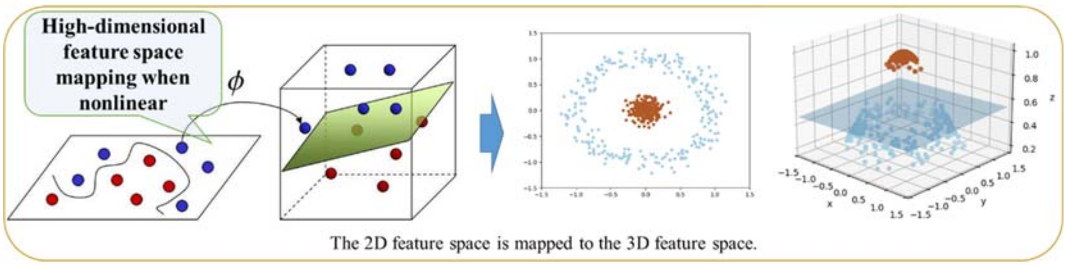


Figure 2-7. The 2D feature space is mapped to the 3D feature space.

2.4 Convolutional Neural Networks

The CNN is a kind of neural network mainly applied in the field of image recognition, which can effectively reduce the dimensionality of large data volumes into small data volume and can effectively retain the image features in accordance with the principle of image processing. The whole network consists of convolutional layers, pooling and fully connection layers.

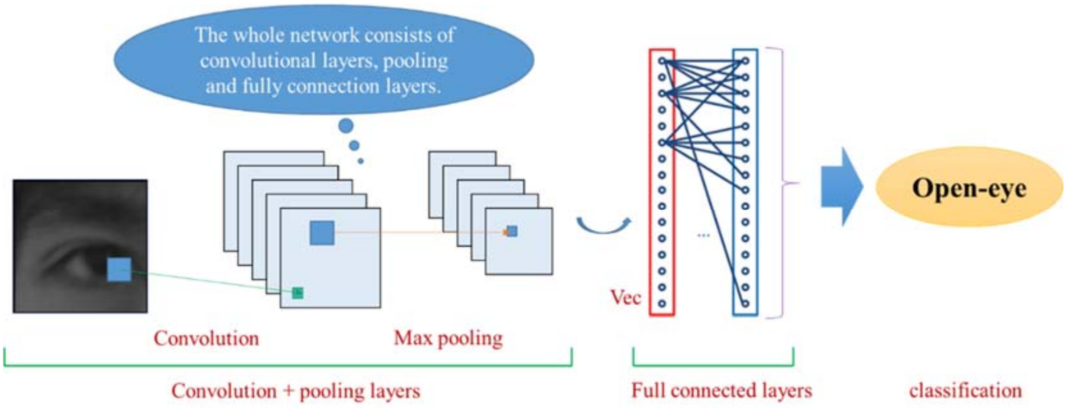


Figure 2-8. Convolutional neural network principle schematic

2.4.1 Interpret each layer calculation of CNN

The convolutional layers are a set of parallel feature maps, which are formed by sliding different convolutional kernels over the input image and performing operations such as those in Figure 2-8a. In addition, at each sliding position, an element-wise multiplication and summation operation is performed between the convolution kernels and the input image to project the information in the perceptual field onto an element of the feature map. This sliding process can be referred to as step size, which is a factor controlling the size of the output feature map. The size of the convolution kernel is much smaller than the input image and overlaps or acts in parallel to the input image. All elements of a feature map are calculated from a single convolution kernel, i.e. a feature map shares the same weights and bias terms.

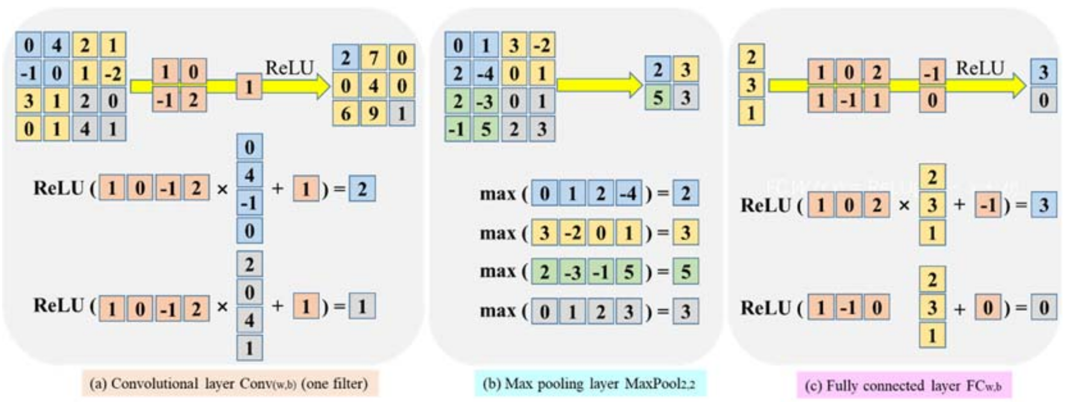


Figure 2-9. One computational example of one of the three layer types of convolutional neural networks.

Pooling is another important concept in convolutional neural networks, and is actually a non-linear form of downsampling. There are several different forms of non-linear pooling functions, of which Max Pooling is the most common. It involves dividing the input image into a number of rectangular regions and outputting the maximum value for each subregion. The pooling layer calculates the output on one pooling window at a time and then moves the pooling window according to the step size. Figure 2-8b below shows the most common pooling layer in use today, with a step size of 2 and a pooling window of 2 x 2 for a 2-dimensional maximum pooling layer. The 2 x 2 blocks are divided from the image every 2 elements and then the maximum value is taken for the 4 numbers in each block. This will reduce the amount of data by 75%.

The fully connected layer acts as a classifier in the entire convolutional neural network. The core operation of full connectivity is to connect each neuron in one layer to each neuron in another layer through a matrix vector product like that in Figure 2-8c. In order to improve the performance of CNN networks, the excitation function for each neuron in the fully connected layer is generally a ReLU function. The output value of the final fully connected layer is passed to an output that can be classified using softmax regression, which can also be called the softmax layer. For a specific classification task, it is important to choose a suitable loss function. In general, the fully-connected layer of a CNN is the same as the MLP structure, and the training algorithm of a CNN also uses the BP algorithm.

2.4.2 Activation Functions

Since the distribution of data is overwhelmingly nonlinear, while the computation of neural networks in general is linear. In convolutional neural networks, the introduction of an activation function is introducing nonlinearity into the neural network and reinforcing the learning ability of the network.

(1) ReLu Function:

$$\varphi(x) = max(0, x) \tag{2-3}$$

ReLu is one of the most common activation functions used as a modified linear unit in artificial neural networks. ReLu function causes values greater than or equal to 0 to be output as is, and values less than or equal to 0 to be set to 0. Its output graph is shown in Figure 2-10.

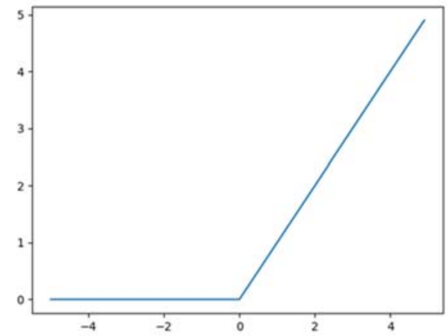


Figure 2-10. Output graph of the ReLu function

(2) Sigmoid Function:

$$\varphi(x) = \frac{1}{1 + e^{-x}} \tag{2-4}$$

The sigmoid function, also called the logistic function, is used for the output of the hidden layer neurons and takes values in the range of (0,1), which can map a real number to the interval of (0,1) and can be used for binary classification. It works well when the difference of features is complex or not particularly large. Its output graph is shown in Figure 2-11.

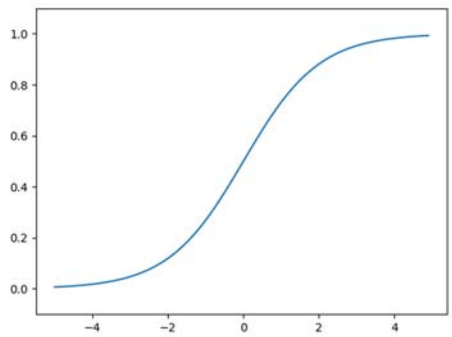


Figure 2-11. Output graph of the Sigmoid function

(3) SoftMax Function:

$$y_i = \frac{\exp(a_i)}{\sum_j^D \exp(a_j)}$$

(2-5)

The Softmax function, as a normalized exponential function, is actually a gradient logarithmic normalization of a finite term discrete probability distribution. Therefore, the Softmax function is widely used in many probability-based approaches to multi-classification problems, including multinomial logistic regression, multinomial linear discriminant analysis, plain Bayesian classifier, and artificial neural networks. In this study, the Softmax function is used in the output layer, where the output of probability is obtained by dividing the output by the sum of all outputs, and the exponential function exp is used to convert to a positive value. Its output graph is shown in Figure 2-12.

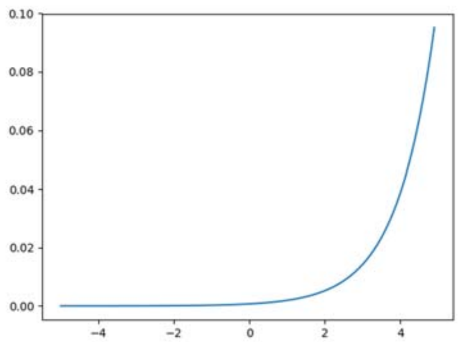


Figure 2-12. Output graph of the SoftMax function

2.5. Simultaneous localization and mapping

The problem of simultaneous localization and mapping (SLAM) can be described as: the robot moves from an unknown position in an unknown environment, localizes itself based on the position and map during the movement, and builds incremental maps based on its own localization to achieve autonomous localization and navigation of the robot [77-82].

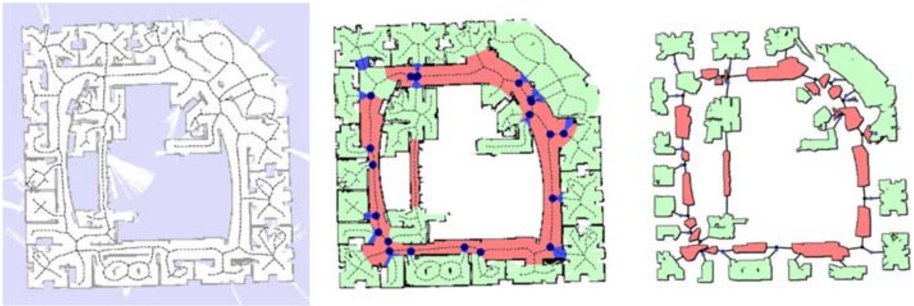


Figure 2-13. Examples of an occupancy grid map, a topological map, and a semantic map of Intel Labs: (left) an occupancy grid map overlaid with Voronoi graph; (middle) a semantic map in which different places in the map are labeled with different colors: hallways in red, rooms in green, doorways in light blue, and junctions in dark blue; (right) a hybrid topological-metric map [79]

SLAM technology covers a very wide range, and there are many ways to classify SLAM according to different sensors, application scenarios, and core algorithms. According to the different sensors, it can be classified into 2D/3D SLAM based on LiDAR, RGBD SLAM based on depth camera, visual SLAM (vSLAM) based on vision sensor, and visual inertial odometry (VIO) based on vision sensor and inertial sensor.

The 2D SLAM based on LiDAR is relatively mature, as early as 2005, Sebastian Thrun et al.'s classic book "Probabilistic Robotics" studied and summarized the 2D SLAM very thoroughly, which basically defined the framework of LiDAR SLAM. In 2016, Google open-sourced the LiDAR SLAM program Cartographer, which can fuse IMU information and unify 2D and 3D SLAM. Currently, 2D SLAM has been successfully used in floor sweeping robots.

Figure 2-14. A 3D map developed by laser ranger measurements using GraphSLAM. Image courtesy of Keiji Nagatani et al [79]

RGBD SLAM based on depth cameras has also developed rapidly in the past few years. Since the introduction of Microsoft's Kinect, there has been a wave of research on RGBD SLAM, and several important algorithms have emerged in just a few years, such as KinectFusion, Kintinuous, Voxel Hashing, DynamicFusion, and so on. Microsoft's Hololens should be integrated with RGBD SLAM, which can achieve very good results in situations where depth sensors can work.

Vision sensors include monocular cameras, binocular cameras, fisheye cameras, etc. Since vision sensors are cheap and can be used both indoors and outdoors, vSLAM is a big hot spot for research [83]. The early vSLAM such as MonoSLAM is more of a continuation of the filtering method in the field of robotics. Nowadays, more optimization methods are used in the field of computer vision, specifically, bundle adjustment in the structure-from-motion method. In vSLAM, according to the extraction method of visual features, it can be divided into feature method and direct method. The current representative algorithms of vSLAM include ORB-SLAM, SVO, DSO, etc.

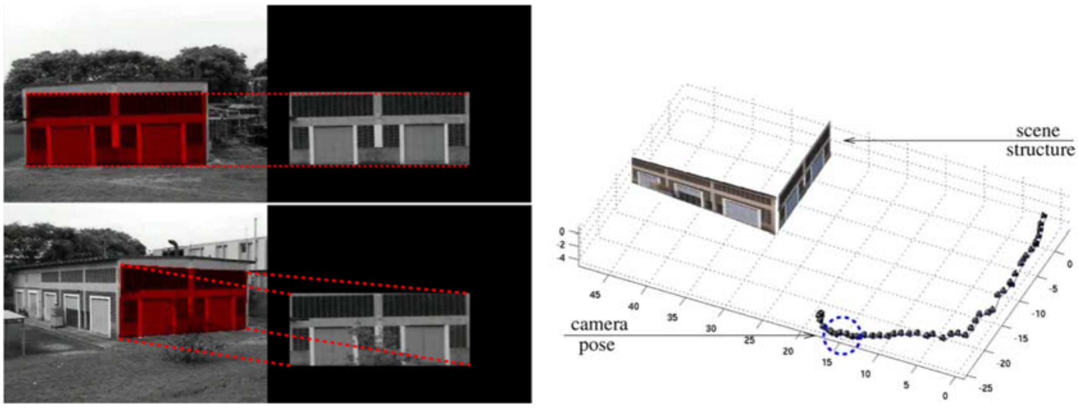


Figure 2-15. HANGAR sequence: 751-frame example of visual SLAM by aligning reference regions to successive images [83]

Vision sensors do not work for texture-free areas. The inertial measurement unit (IMU) can measure the angular velocity and acceleration through the built-in gyroscope and accelerometer, and then project the camera pose, but the projected pose has accumulated errors. Since vision sensors and IMUs are very complementary, VIOs that fuse their measurement information are also a hot research topic. VIO can be divided into filter-based methods and optimization-based methods according to the different information fusion methods, and the representative algorithms of VIO include extended kalman filter (EKF), Multi-State Constraint Kalman Filter (MSCKF), pre-integration, Open Keyframe-based Visual-Inertial SLAM (OKVIS), etc. [84-87].

Figure 2-16. The global map and round-trip trajectory overlaid on the Google satellite map in an fully autonomous flight experiment. The blue, red, and yellow dots represents the starting point, goal location, and the only entrance of the warehouse respectively. The global laser point cloud is registered using the estimation produced by the S-MSCKF. Over 700 m trajectory, the final drift is around 3 m under 0.5% of the total traveled distance [87]

2.6 Vehicle to everything

Vehicle to everything (v2x: Vehicle to X) is the exchange of information from the vehicle to the outside world. By integrating global positioning system (GPS) navigation technology, vehicle-to-vehicle communication technology, wireless communication and remote sensing technology, Telematics has laid down a new direction for the development of automotive technology and realized the compatibility of manual driving and automatic driving [88-90].

V2X is a key technology for future intelligent transportation systems. It enables communication between vehicles and vehicles, vehicles and base stations, and base stations and base stations. Thus, a series of traffic information such as real-time road conditions, road information, pedestrian information, etc. can be obtained to improve driving safety, reduce congestion, improve traffic efficiency, provide in-car entertainment information, etc.

In simple terms, models equipped with the system can automatically select the best route for road conditions by analyzing real-time traffic information in autonomous driving mode, thereby greatly reducing traffic congestion. In addition, through the use of on-board sensors and camera systems, it can also sense the surrounding environment and make rapid adjustments, thus achieving "zero traffic accidents". For example, if a pedestrian suddenly appears, it can automatically slow down to a safe speed or stop.

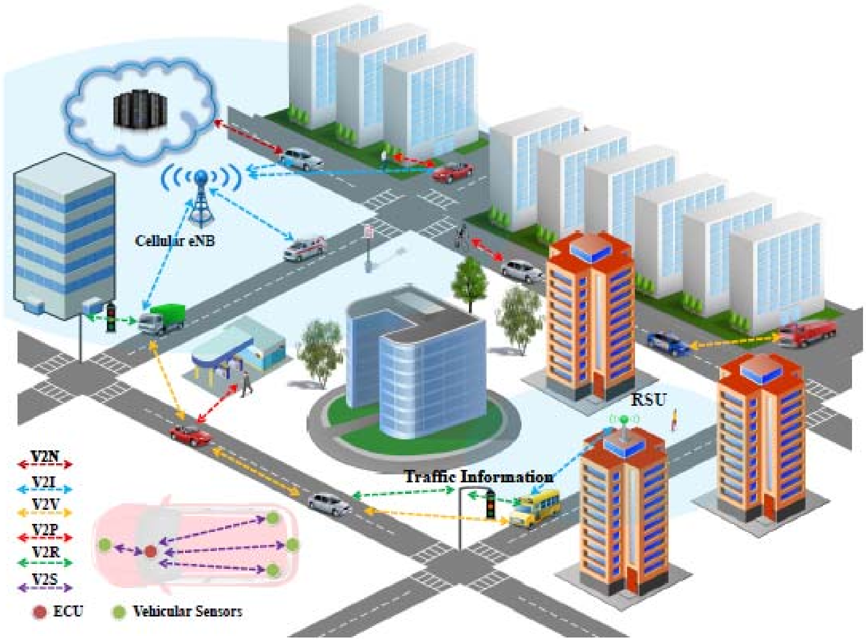


Figure 2-16. An overview of V2X scenario [88]

Chapter 3. A human-computer control system based on intelligent recognition of eye movements and its application in wheelchair driving

This paper presents a practical human-computer interaction system for wheelchair motion through eye tracking and eye blink detection. In this system, the pupil in the eye image has been extracted after binarization, and the center of the pupil was localized to capture the trajectory of eye movement and determine the direction of eye gaze. Meanwhile, convolutional neural net-works for feature extraction and classification of open-eye and closed-eye images have been built, and machine learning was performed by extracting features from multiple individual images of open-eye and closed-eye states for input to the system. As an application of this human- computer interaction control system, experimental validation was carried out on a modified wheelchair and the proposed method proved to be effective and reliable based on the experimental results.

3.1 Introduction

Human-computer interaction (HCI) has been widely studied since the 1960s with the rapid development of information systems, which aims to design a human-computer interface with ergonomic characteristics [1]. HCI systems in automated devices have been based on the traditional interface with the monitor, keyboard, and mouse for a long time. However, this manual input HCI was cumbersome to use, and to change this situation, HCI with gesture-controlled interfaces has been widely studied [2–4]. Nevertheless, there are many physically disabled people in real life who still are unable to use these devices or even to travel independently. These physically disabled people are completely dependent on others for their daily needs [5]. In order to improve the quality of life for people with disabilities, HCI systems without relying on hands and feet is particularly important.

In recent years, there has been much research on control systems for people with ALS (amyotrophic lateral sclerosis) and other severe physical disabilities that use biological information such as eye movements, EEG, and EMG directly for HCI, without the use of hands and feet [6–12]. As some particular interest research works are studies on eye gaze input and eye blink, etc., which transmit information through the human eye. Because the signal conveyed by the eye has greater stability and real-time compared to EEG and EMG, humans can communicate a great deal of information more quickly and directly through the eye [13–25]. Additionally, the problem of calibration in eye gaze regarding the difficulty recognition was obtained due to the method using smooth pursuit motion does not require calibration, which is of interest for eye gaze input studies because it makes the calibration of gaze input easy [21].

This paper presents a practical HCI system for wheelchair control. Eye tracking and eye blink detection based on image processing techniques have been integrated into the proposed system. Since subjects can convey incorrect information when they are slightly inattentive during eye gaze interactions, eye blink detection was introduced to assist in completing this study. In contrast to some commonly used methods [22–25], image processing techniques were applied as a new approach, which improved the reliability and convenience of the system. After the captured image was processed by image processing, its features were extracted, which made it easier to be recognized and the accuracy of recognition becomes higher. Similarly, the eye object detection module in the Haar Cascade (a Face Recognition Module) was used in the program, and as soon as the system turned on, it was able to quickly capture and locate the eye region from the images captured by the camera.

In this study, for eye gaze detection, pupil features were extracted from the eye images using image binarization. The eyeball movement trajectory was tracked by locating the pupil, based on the eyeball movement trajectory, the subject's eye gaze area was deter-mined so that the information that the subject wants to express through eye gaze can be obtained. For eye blink detection, some machine learning methods have been used to accomplish eye blink detection in order to solve the problems of low accuracy, lack of stability, and inconvenience of use of other methods. Machine learning techniques were used very successfully in the field of computer vision, where it was used to simulate human intelligence by learning the surrounding environment [26]. Machine learning was used in this study to extract image features of the open and closed states of multiple individual eyes to obtain a learning model so that the system could discriminate whether a subject the eye blink or not by take real-time eye images of the subject. Three methods have been used to determine the eye blink: pixel ratio [27], support vector machine (SVM) [28], and convolutional neural network (CNN) [29] to determine eye blink. In comparison, using CNN had the highest accuracy in detecting eye blink.

The HCI system of this study was mounted on a modified electric wheelchair for experimental validation. Experiments with many methods to drive wheelchairs for disabled people have been used in many HCI systematic studies, which gives us some references [30–36]. Based on the data obtained in the experiments, it can be demonstrated that this integrated multi-domain interaction system is effective. This also provides some basis for further research on gaze interaction and HCI.

3.2 HCI Control System of Wheelchairs

3.2.1 System Overview

The HCI system of this study was applied to control the movement of an electric wheelchair (see Figure 3-1). The testing system included a blink detection device, head-mounted eye gaze tracking device (see Figure 3-1). Wireless communication was used between the interactive system and the wheelchair, and the subject used eye gaze direction and eye blink to control the wheelchair movement. Throughout the process of controlling the movement of the wheelchair, the subject interacts with the computer, communicating information to the computer through the eyes, and the computer communicates commands to the wheelchair drive unit to control the wheelchair movement.

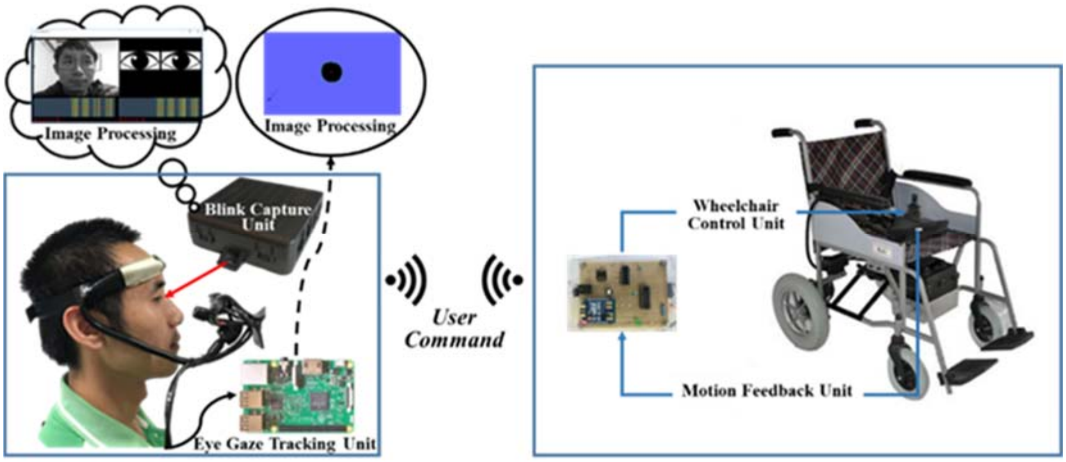


Figure 3-1. Schematic diagram of the HCI system for eye state recognition.

3.2.2 Hardware Systems for Wheelchairs

In this study, an aluminum AR-200 wheelchair manufactured by Matsunaga Seisakusho Ltd. was used. The hardware part of the wheelchair was improved and we designed a drive unit for the wheelchair. This allows us to control the travel of the electric wheelchair by eye gaze in the direction or by eye blink. In the modification, the rocker control unit of the wheelchair was still retained so that the wheelchair could still be controlled by the rocker when switching to manual operation. Figure 3-2 shows the hardware of the improved power wheelchair.

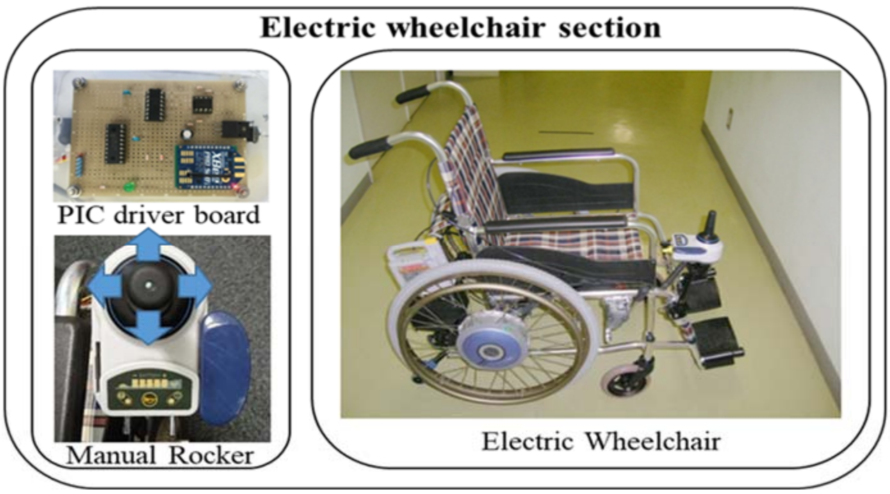


Figure 3-2. Modified electric wheelchair hardware display image.

In order to be able to drive the wheelchair through the interactive system, a driver board was designed and developed for the wheelchair. Figure 3-3 shows the circuit schematic of the drive circuit board. This is very important as it makes our interactive system highly portable without having to consider the complex communication protocols of the controlled devices. A wireless communication serial port has been used in the driver board so that wireless communication could be made between the computer and the wheelchair.

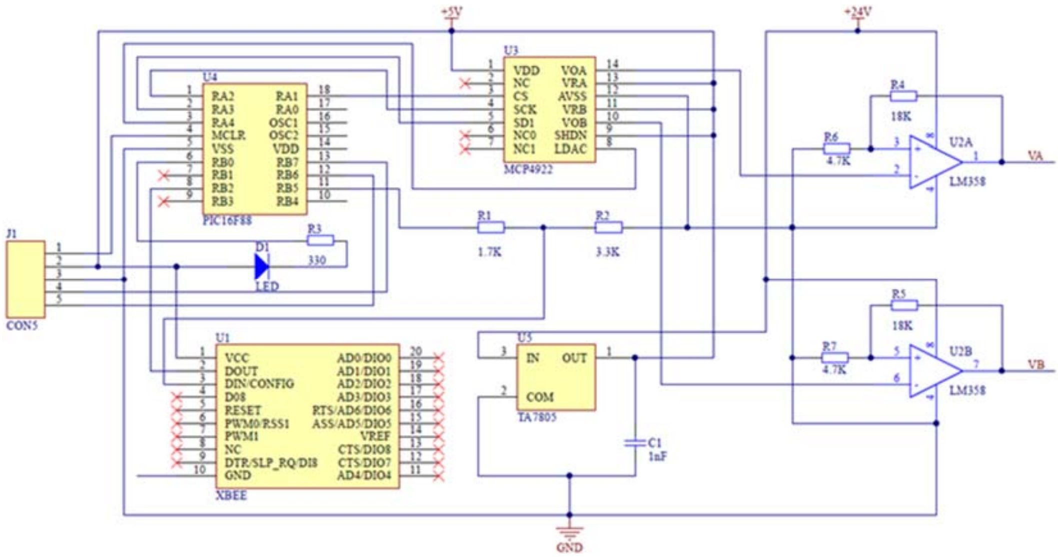


Figure 3-3. Circuit schematic of the driver circuit board.

3.3. Eye Movement Recognition Methods

3.3.1. Eye Gazes Detection Method

The eye gaze tracking device (see Figure 3-4) used contrast to locate the center of the pupil and used infrared non-columnar light to generate corneal reflections, capturing the black area of the pupil by illuminating the eye with infrared light [17]. In order to be able to use in a variety of light environments like sunlight, a filter was added to the lens, which is to be able to filter out light other than infrared light.

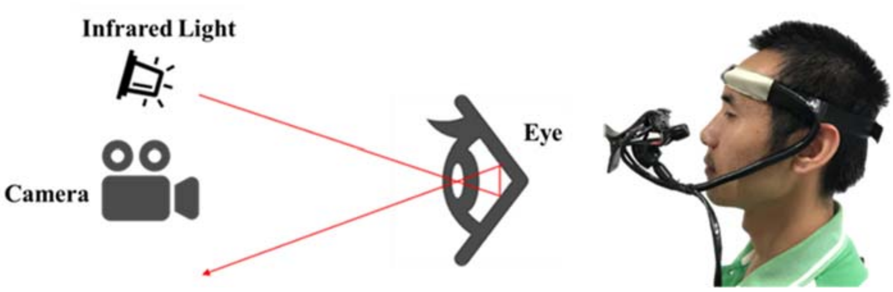


Figure 3-4. Diagram of corneal reflection eye tracking technology.

The image binarization method was used to process the captured eye images. Image binarization was the process of converting an image with shadows into two shades of black and white. A threshold was preset, and if the value of each pixel was above the threshold, it was replaced with white, and if it was below the threshold, it was replaced with black.

$$f'(x,y) = \begin{cases} 255 & f(x,y) \geq \theta \\ 0 & f(x,y) < \theta \end{cases}$$

(3-1)

Here, the threshold is θ, the coordinate system is (x, y), the concentration value is $f(x, y)$, and the transformed concentration value is $f'(x, y)$. The flow of the eye tracking technique is shown in Figure 3-5. The pupil in the eye image was extracted by binarization, and the pupil center coordinates were calculated. The gaze direction can be computed based on the motion trajectory of the pupil center coordinates.

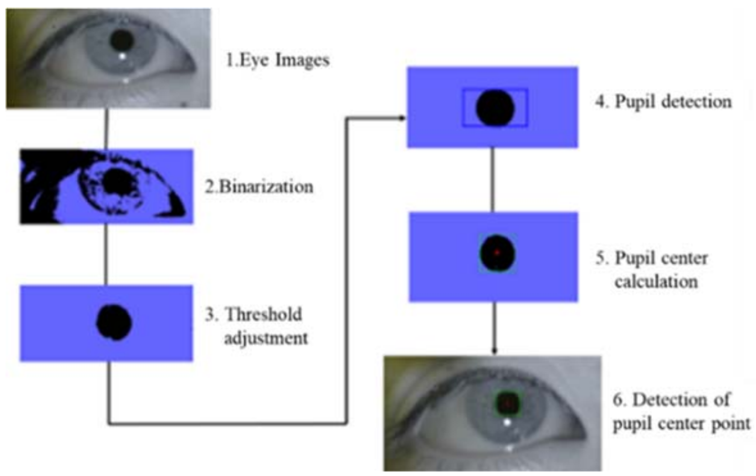


Figure 3-5. Flow chart of eyeball tracking technology.

The image of the eye's moveable range was divided into five regions, and the five regions were labeled 2, 4, 5, 6, and 8 according to the numeric keypad. These 5 regions can be used to represent different commands for different driving devices, and of course more regions can be delineated depending on the driving device, so that more commands can be obtained. In this article, when controlling the wheelchair movement, the commands for these 5 zones are: backward, left turn, stop, right turn and forward. The five areas set up are shown in the Figure 3-6.

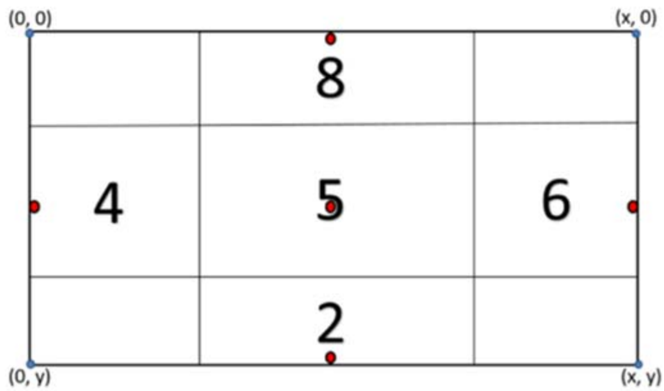


Figure 3-6. The command chart corresponding to the eye gaze area

3.3.2. Eye Blink Detection Method

3.3.2.1. Pixel Ratio

The value of each pixel in the image is represented by 8-bit unsigned characters (value range: 0–255), which can be converted to 0 values by binarization if it is below the threshold, or to the highest value of 255 if it is above the threshold. The image is shown in Figure 3-7a shows a black-and-white image of the open-eye state and Figure 3-7b shows a black-and-white image of the closed-eye state. By comparing the two images below it can be seen that the binarised image varies according to the open/closed state of the eye. Thus, if different values are obtained from the two images, the images can be discriminated.

(a) **(b)**

Figure 3-7. Binarized black and white images of the eye with the eye open and closed. (a) Open eye binary image; (b) Close eye binary image.

First, acquire the following eye image with a resolution of 64 × 64 pixels. The number of pixels will be 4096. Next, count the number of black pixels in the image. If there are 1024 pixels, the percentage of black pixels is (1024/4096) × 100, or 25%. This ratio was higher when the eye is open and lower when the eye is closed, so when the eye blink, this ratio changes. Figure 3-8 shows the Data waveform of eye blink image pixel ratio change.

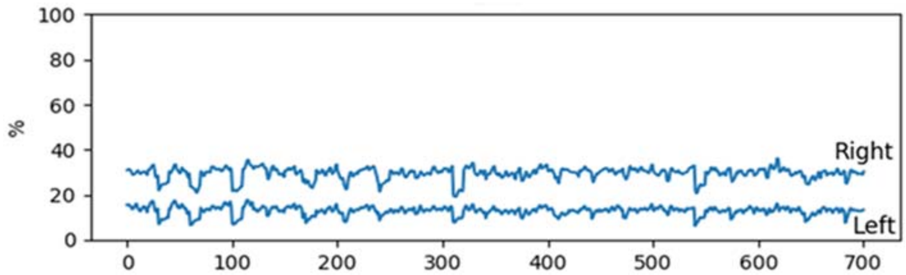


Figure 3-8. Data waveform of eye blink image pixel ratio change.

3.3.2.2. Support Vector Machine

One of the oldest methods used in image classification is the SVM [28]. It is one of the pattern recognition models that uses supervised learning. It differs from ordinary pattern recognition models in that it performs margin maximization and kernel tricks.

1. Margin maximization

Margin maximization is the shortest distance between the boundary and the data. Figure 3-9 shows the image of marginal maximization, the idea of margin maximization is to draw the boundary as far away as possible from the data that is closest to the boundary between the two classes.

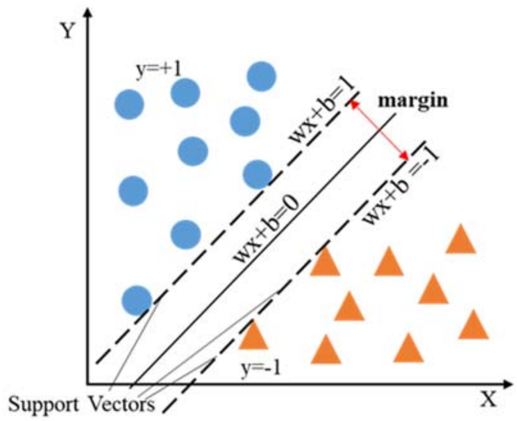


Figure 3-9. Schematic diagram of margin maximization.

2. Mapping of feature space

In encountering nonlinear separation, it is necessary to use kernel tricks in SVM to map the data from the original space to the new space (see Figure 3-10, mapping 2-dimensional feature space to 3-dimensional feature space), and then the training data are used in the new space to get the learning model using linear methods.

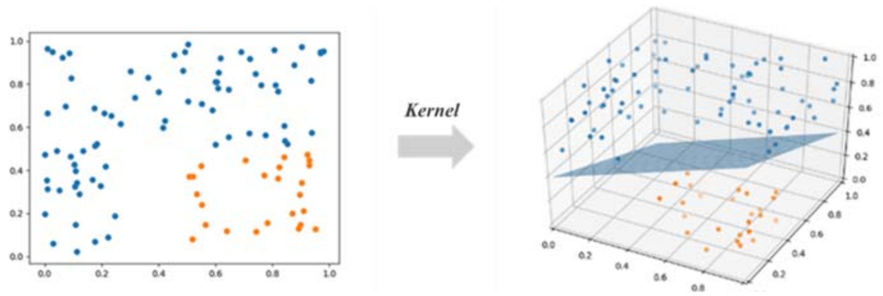


Figure 3-10. The 2D feature space is mapped to the 3D feature space.

To find a map to feature space, we need to find the inner product $\varphi(x)^T\varphi(y)$on the feature space. The kernel trick allows us to calculate the inner product on the feature space without knowing what the feature space is and what φ is.

Discriminant function:

$$f(x) \quad = sign[\sum_{i=1}^{n} a_i\, y_i \varphi(x_i)^T \varphi(x)] \tag{3-2}$$

(a: Weight; y: Label (1 or -1); x_i: *i-th* learning data; x: Input data)

Kernel function:

$$k(x, y) = \varphi(x)^T \varphi(y)$$

(3-3)

(Inner product in feature space: $\varphi(x)^T \varphi(y)$)

Replace the inner product of the discriminant function:

$$f(x) \quad = sign\left[\sum_{i=1}^{n} a_i \, y_i k(x_i, x)\right]$$

(3-4)

3.3.2.3. Convolutional Neural Network

The Convolutional Neural Network (CNN) is an important method in the field related to pattern recognition [29]. This research constructed a learning model with CNN and used it to detect blinks of the subject. Figure 3-11 shows the CNN model constructed in this study. In the constructed CNN model input images of open and closed eyes with a resolution of 64×64 were subjected to convolutional operations, the feature values of the input images were extracted, and a recognition model was built based on these feature values and it was applied to the system.

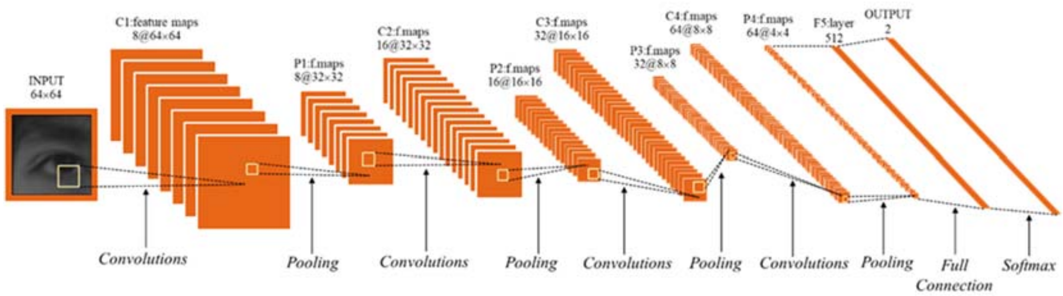


Figure 3-11. The structure of CNN that has been built.

In this experiment, the well-known VGG-16 discriminative model has been used as a reference. Since only two types of images (open-eye and closed-eye) need to be feature extracted and classified, the model was finally reduced from the original 16 layers to 10 layers and the convolutional layers from 13 layers to 8 layers after continuous experiments in order to balance high efficiency and high accuracy. The relevant parameters of each layer are shown in Table 1.

Table 1. Relevant parameters of each layer.

Layer Name	Layer Type	Relate Parameters
Conv1_1	convolution	3 × 3, 8, relu, stride1
Conv1_2	convolution	3 × 3, 8, relu, stride1
Pool1	Pooling	2 × 2, 8, max pool, stride2
Conv2_1	convolution	3 × 3, 16, relu, stride1
Conv2_2	convolution	3 × 3, 16, relu, stride1
Pool2	Pooling	2 × 2, 16, max pool, stride2
Conv3_1	convolution	3 × 3, 32, relu, stride1
Conv3_2	convolution	3 × 3, 32, relu, stride1
Pool3	Pooling	2 × 2, 32, max pool, stride2
Conv4_1	convolution	3 × 3, 64, relu, stride1
Conv4_2	convolution	3 × 3,64, relu, stride1
Pool4	Pooling	2 × 2, 64, max pool, stride2
Fuc1	Fully-connected	512, sigmoid
Drop	Dropout	dropout-ratio 0.5
Fuc2	Fully-connected	2, softmax

3.3.2.4. The Eye Blink Detection Device and Its GUI

The graphical user interface of the eye blink detection system is shown in Figure 3-12, which displays the camera image, the eye status picture area, the eye blink waveform, and the number of consecutive eye blinks by the user. In the upper left area, the camera image was displayed. The eye status picture was displayed in a rectangular area, and the position of the eye area was updated every 10 frames. The two waveforms at the bottom represent the open and closed state of the eye, 1 when the eye is open and 0 when it is closed. The image at the top right shows the state of the eye in the form of an image. The text string at the bottom indicates the number of consecutive eye blinks. The symbol (*) increases with each blink of the eye. The face and eyes object detection module in a face recognition module (Haar Cascade) was used in the program in order to quickly capture the face and locate the eye area, so that the capture frame in the screen always firmly captures both eyes. During the control of the wheelchair, the wheelchair moves forward after the subject eye blinks three times rapidly. After the subject blink once with the right eye, the wheelchair turns right. After the subject blink left once, the wheelchair turns left. After the subject eye blinks four times quickly, the wheelchair moves backward. After the subject eye blinks twice quickly, the wheelchair stops moving.

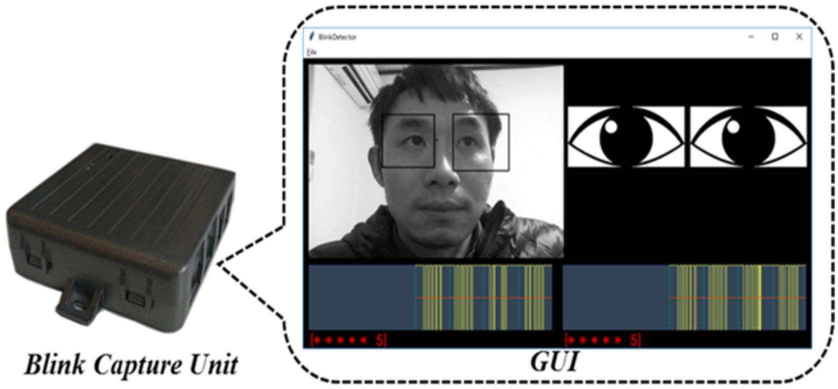


Figure 3-12. Diagram of blink detection device and its graphical user interface. The red symbol (*) in the interface increases with each blink of the eye.

3.4 Results and discussion

3.4.1. Results of Eye Gaze Direction Recognition Experiments

Figure 3-13 Comparison of eye images in various states during the experiment. Figure 3-13a shows the eye image without infrared light irradiation, and Figure 3-13b shows the image of the eye under infrared illumination. A comparison between the two can be found in Figure 3-13b, where the pupil contour was more well-defined. Because corneal reflection is a way to detect pupils darker, the iris colours can be compared by separating the infrared light from the optical axis of the eye-tracking camera. Figure 3-13c shows the eye image after the binarization process. To highlight the binarization changes, the background color was set to blue instead of white. After binarization there were only two colors left in the plot, the next step was to adjust the threshold to make the pupil outline more prominent and obvious in the image.

(a) (b) (c)

Figure 3-13. Comparison of eye images during the experiment. (a) The eye image without infrared light irradiation; (b) The image of the eye under infrared illumination; (c) The eye image after the binarization process.

3.4.2. Eye Blinks Recognition Experimental Results

In this study, three methods, pixel ratio, SVM, and CNN, were used to determine blinking, and experimental comparisons were made separately.

1. Pixel Ratio:

In this experiment, a 5% change in the pixel ratio was used as an eye blink condition. In addition, some changes can be obtained from the image four frames ago, since the changes were not significant compared to the image one frame ago. Figure 3-14 shows the judgment waveforms of eye blink detection by the pixel ratio method in this study. During the experiment, a total of 20 blinks was made, only two blinks were detected in the left eye and 10 blinks were detected in the right eye.

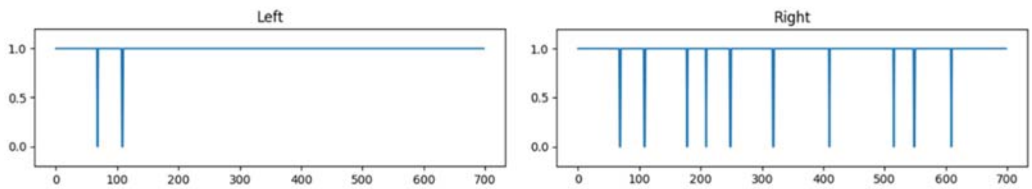


Figure 3-14. Judgment waveforms for eye blink detection by pixel ratio. Left: Blink judgement waveform in the left eye; Right: Blink judgement waveform in the right eye.

2. Support Vector Machine:

In the experiment using the SVM approach to detect eye blink, the testers blinked both eyes the same 20 times. Figure 3-15 shows judgment waveforms for eye blink detection by SVM, and we can see that 15 blinks were detected in the left eye and 19 blinks were detected in the right eye.

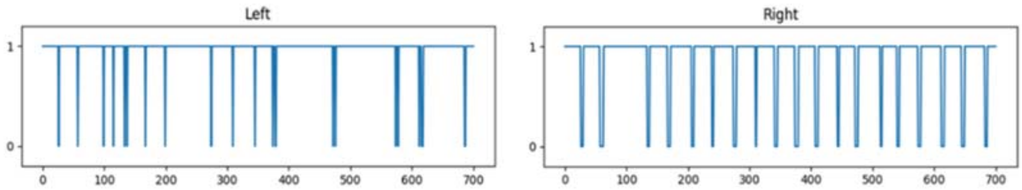


Figure 3-15. Judgment waveforms for eye blink detection by SVM. Left: Blink judgement waveform in the left eye; Right: Blink judgement waveform in the right eye.

3. Convolutional Neural Network:

Convolution layers can combine different local structures to present more useful features in a region. By convolving the input image, the input image is rendered with more feature maps. Figure 3-16 shows the convolutional layer feature map data of this study in the experiment. In the first and second layers, there does not seem to be much change. However, in the third layer, the brightness becomes more diversified, indicating that the network can adapt to changes in the brightness of the input image. In the fourth layer, the extracted features were different from those in the previous layers, and the contours were emphasized. Five to six layers show that the contours were greatly emphasized. In the seventh layer, more features were extracted. In the eighth and final layer, the features were extracted to the extent that the original form was no longer visible to the eye.

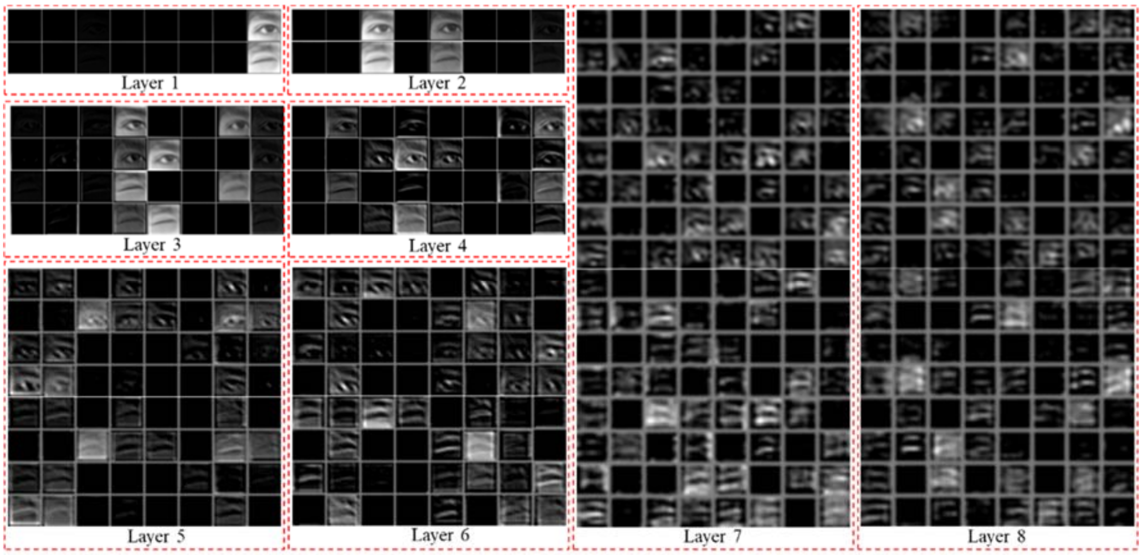


Figure 3-16. The feature map of convolution layer.

Figure 3-17 shows the judgment waveforms of eye blink detection experiment with CNN.

The testers performed 20 blinks in both eyes. Based on the data waveform plot it can be seen that all 20 blinks in the left and right eyes were detected by the system.

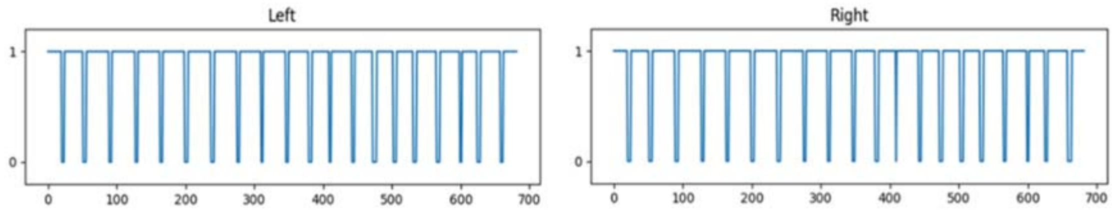


Figure 3-17. Judgment waveforms for eye blink detection of CNN. Left: Blink judgement waveform in the left eye; Right: Blink judgement waveform in the right eye.

4. Comparative discussion of results

One hundred experiments were done for each of the three methods of identifying blinks. Table 2 shows the results of the number of detections for 100 blinks, and Figure 3-18 shows a comparison between the experimental data of the three detection methods. According to the experimental results, the recognition rate using CNN is the highest with 99% accuracy. This indicates that CNN method is still optimal, and the next step is just to adjust each parameter to further improve the accuracy rate. In comparison, the Pixel ratio method is the simplest of the three methods, which does not require the acquisition of human eye images of the subject for learning and runs relatively fast, however, the detection rate is the worst. Although a neural network was built in the SVM method, however it also does not perform image convolution and has the second fastest running speed and has the second lowest detection rate. Finally, the convolutional neural network has the highest detection rate among the three methods, and it is also the most complex of the three methods. In operation, subjects can use different methods depending on the conditions and are able to collect a wider range of data.

Table 2. Number of detections per 100 blinks.

Methods	Detection Count	Undetected Count	False Positives Count
pixel ratio	28	72	0
SVM	74	26	11
CNN	99	1	0

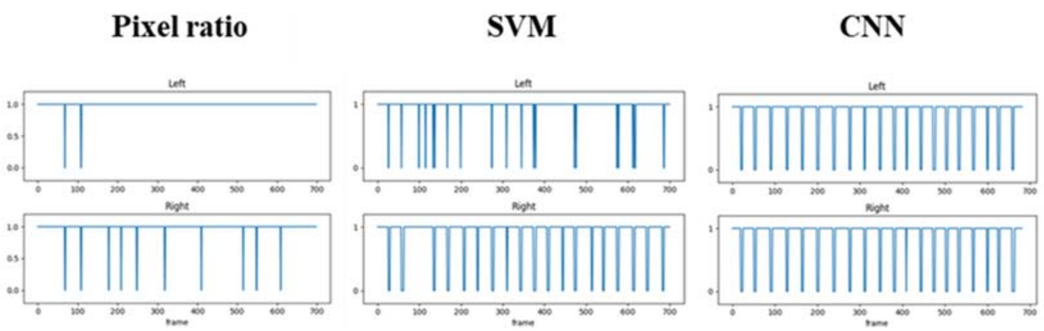


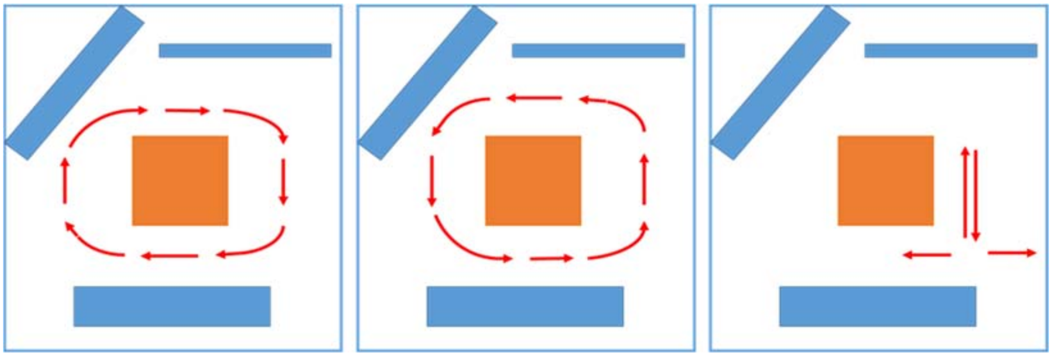
Figure 3-18. Comparison among the experimental data of the three detection methods. Pixel ratio: Blink judgement waveform using Pixel ratio; SVM: Blink judgement waveform using SVM; CNN: Blink judgement waveform using CNN.

3.4.3. Drive Experiment Results and Discussion

In this study, experiments were conducted using an eye-movement controlled wheelchair and the results are discussed.

1. Drive experiment results

Figure 3-19 shows the roadmap for driving the wheelchair movement during the experience, and some obstacles were set up in the test site in order to increase the difficulty. Twelve subjects controlled the wheelchair movement by eye gaze and eye blink according to the path plan in Figure 3-19 and each experimented once. All 12 subjects successfully completed the experiment and were able to control their wheelchairs very smoothly to avoid obstacles during the experiment. The shortest time was 2 min and the longest time was 3 min for all 12 subjects to complete the experiment. Figure 3-19a shows the subjects driving the wheelchair clockwise, and after reaching that end point, the subjects turned around in the same place and moved counterclockwise according to the route in Figure 3-19b, and completed the whole process after reaching the end point. Figure 3-19c shows the subject's remotely controlled wheelchair moving according to the route in the figure. The subject first controls the wheelchair straight ahead to reach the preset point, and then controls the wheelchair backward to the starting point, the subject then controls the wheelchair to turn 90° left, and then 180° right to complete the process.



(a) (b) (c)

Figure 3-19. Drive wheelchair movement route planning map. (a) Clockwise driving route; (b) Counter-clockwise driving route; (c) Remote control of driving routes.

Figure 3-20 shows a combined image of experimental data using eye gaze to drive the wheelchair movement, where subjects controlled the wheelchair to move along the route planned in Figure 3-19(a) (b). The subject sat in the wheelchair and controlled the wheelchair to move clockwise from the starting point by eye gaze. When the controlled wheelchair reachs the preset end point, it was controlled to stop, turn around in place, and continue to move counterclockwise to the preset end point, thus ending the process.

Figure 3-20. Eye gaze direction control wheelchair movement experiment.

In the experiment using eye blink controlled wheelchair movement, in order to differentiate the experiments using eye gaze controlled wheelchair movement, the subject sat in front of a monitor with the eye blink detection device placed in front of the eyes, and remotely controlled the wheelchair to move along the route planned in Figure 3-19c. Figure 3-21 shows a data set plots of eye blink controlled wheelchair movement experiment. In the figure set, the first row of images (number: 1–4) shows the blink controlled wheelchair moving forward, the second row of images (number: 5–8) shows the blink controlled wheelchair moving backward to its original position, the third row of images (number: 9–12) shows the blink controlled wheelchair turning 90° to turn left, and the fourth row of images (number: 13–16) shows the blink controlled wheelchair turning 180° to turn right. This concludes an experimental cycle.

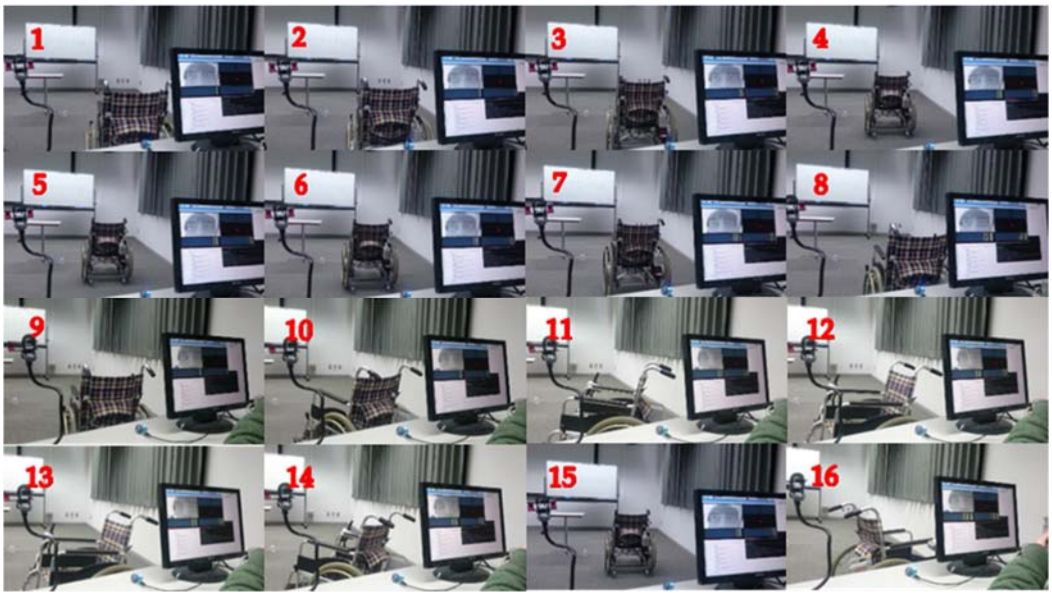


Figure 3-21. Eye blink control wheelchair travel experiment.

2. Discussion of results

The experimental results show that gaze and eye blink could effectively control the wheelchair to complete the related movements. However, the test subjects had to concentrate on controlling the wheelchair movement throughout the experiment, which was a problem. As a future research topic, the controlled device can provide some feedback to the human body so that it can be alerted in time when the operator is not focused enough.

3.5 Conclusions

In this paper, the extraction of eye movement information to control relevant mechanical movements was investigated. An eye-movement based HCI system that communicates information to the machine through the human eye was developed in this research. The binarization of the image was used to localize the pupil in this study. Since the difference between the two colors in the eye was apparent, binarization of the pupil image enables the localization of the black pupil, and based on the localization of the pupil, the direction of eye gaze could be determined. Similarly, pixel ratio modelling has been fully used in the eye blink detection. Through operational experiments, it could be confirmed that the eye blink state could be detected, however, the recognition rate could be further improved. Therefore, the machine learning methods SVM and CNN were used for eye blink detection and the accuracy was significantly improved, especially the eye blink detection using CNN method, was tested to reach 99% accuracy.

To further validate these, the system was ported to a power wheelchair. We first modified the hardware of the wheelchair and developed a hardware drive system for the electric wheelchair. This allowed the wheelchair to receive commands from the subject through eye movements. In the experiment, the subject sat in the wheelchair to control its movement, or controls the wheelchair movement remotely. The wheelchair was controlled and followed a set route, successfully avoiding obstacles along the way to the end point, and was tested repeatedly and successfully. The experiments proved that the method of obtaining information expressed by human eyes through image parsing and machine learning has been effective. Future research can optimize the model on this basis to obtain more information conveyed by the human eye, for example, the fatigue level and mental state of a person can be judged by the human eye.

References

[1] Hoc, J.M. From human-machine interaction to human-machine cooperation. Ergonomics 2000, 43, 833–843.

[2] Memo, A.; Zanuttigh, P. Head-mounted gesture controlled interface for human-computer interaction. Multimed. Tools Appl. 2018, 77, 27–53.

[3] Drewes, H.; Schmidt, A. IFIP Conference on Human-Computer Interaction, Interacting

with the Computer Using Gaze Gestures; Springer: Berlin/Heidelberg, Germany, 2007; pp. 475–488.

[4] Pfeuffer, K.; Alexander, J.; Chong, M.K.; Gellersen, H. Gaze-touch: Combining gaze with multi-touch for interaction on the same surface. In Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology; ACM: New York, NY, USA, 2014.

[5] Bickenbach, J.E.; Chatterji, S.; Badley, E.M.; Üstün, T.B. Models of disablement, universalism and the international classifica-tion of impairments, disabilities and handicaps. Soc. Sci. Med. 1999, 48, 1173–1187.

[6] Sorbello, R.; Tramonte, S.; Giardina, M. E; La Bella, V.; Spataro, R.; Allison, B.; Guger, C.; Chella, A. A Human–Humanoid Interaction Through the Use of BCI for Locked-In ALS Patients Using Neuro-Biological Feedback Fusion. IEEE Trans. Neural Syst. Rehabil. Eng. 2018, 26, 487–497.

[7] Rupom, A.I.; Patwary, A.B. P300 Speller Based ALS Detection Using Daubechies Wavelet Transform in Electroencephalo-graph. In Proceedings of the International Conference on Electrical, Computer and Communication Engineering, Cox'sBazar, Bangladesh, 7–9 February 2019.

[8] Korovesis, N.; Kandris, D.; Koulouras, G.; Alexandridis, A. Robot motion control via an EEG-based brain--computer inter-face by using neural networks and alpha brainwaves. Electronics, 2019, 8, 1387.

[9] Deligani, R.J.; Hosni, S.I.; Borgheai, S.B.; McLinden, J.; Zisk, A.H.; Mankodiya, K.; Shahriari, Y. Electrical and Hemodynamic Neural Functions in People With ALS: An EEG-fNIRS Resting-State Study. IEEE Trans. Neural Syst. Rehabil. Eng. 2020, 28, 3129–3139.

[10] Chowdhury, T.T.; Fattah, S.A.; Shahnaz, C. ALS Detection Based on T-Location Scale Statistical Modeling of the DWT Coeffi-cients of EMG Signals. In Proceedings of the IEEE International Conference on Biomedical Engineering, Computer and In-formation Technology for Health (BECITHCON), Dhaka, Bangladesh, 1 December 2019.

[11] Carrasquilla-Batista, A.; Quiros-Espinoza, K.; Gomez-Carrasquilla, C. An Internet of Things (IoT) application to control a wheelchair through EEG signal processing. In Proceedings of the 2017 International Symposium on Wearable Robotics and Rehabilitation

(WeRob), Houston, TX, USA, 5–8 November 2017.

[12] Larson, A.; Herrera, J.; George, K.; Matthews, A. Electrooculography based electronic communication device for individuals with ALS. In Proceedings of the IEEE Sensors Applications Symposium (SAS), Glassboro, NJ, USA, 11–13 March 2019.

[13] Hansen, D.W.; Ji, Q. In the Eye of the Beholder: A Survey of Models for Eyes and Gaze. IEEE Trans. Pattern Anal. Mach. Intell. 2010, 32, 478–500.

[14] T.E. Hutchinson, T.E.; White, K.P.; Martin, W.N.; Reichert, K.C.; Frey, L.A. Human-computer interaction using eye-gaze in-put. IEEE Trans. Syst. Man Cybern. 1989, 19, 1527–1534.

[15] Shinde, S.; Kumar, S.; Johri, P. A Review: Eye Tracking Interface with Embedded System & IOT. In Proceedings of the International Conference on Computing Power and Communication Technologies (GUCON), Greater Noida, India, 28–29 September 2018.

[16] Wanluk, N.; Visitsattapongse, S.; Juhong, A.; Pintavirooj, C. Smart wheelchair based on eye tracking. In Proceedings of the 9th Biomedical Engineering International Conference (BMEiCON), Laung Prabang, Laos, 7–9 December 2016.

[17] Li, X.; Wee, W.G. An efficient method for eye tracking and eye-gazed FOV estimation. In Proceedings of the 16th IEEE In-ternational Conference on Image Processing (ICIP), Cairo, Egypt, 7 November 2009.

[18] Arai, K.; Mardiyanto, R. Eyes Based Eletric WheelChair Control System. Int. J. Adv. Comput. Sci. Appl. 2011, 2, 98–105.

[19] Araujo, J.M.; Zhang, G.; Hansen, J.P.P.; Puthusserypady, S. Exploring Eye-Gaze Wheelchair Control. In ACM Symposium on Eye Tracking Research and Applications; ACM: New York, NY, USA, 2020.

[20] Hansen, J.P.; Trudslev, A.K.; Harild, S.A.; Alapetite, A.; Minakata, K. Providing access to VR through a wheelchair. In Ex-tended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems; ACM: New York, NY, USA, 2019.

[21] Pfeuffer, K.; Vidal, M.; Turner, J.; Bulling, A.; Gellersen, H. Pursuit calibration: Making gaze calibration less tedious and more flexible. In Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology; ACM: New York, NY, USA, 2013.

[22] Purwantoa, D.; Mardiyanto, R.; Arai, K. Electric wheelchair control with gaze direction

and eye blinking. Artif. Life Robot. 2009, 14, 397–400.

[23] Varela, M. Raw eeg signal processing for bci control based on voluntary eye blinks. In Proceedings of the 2015 IEEE Thirty Fifth Central American and Panama Convention, Tegucigalpa, Honduras, 11–13 November 2015.

[24] Gomez-Carrasquilla, C.; Quiros-Espinoza, K.; Carrasquilla-Batista, A. Wheelchair control through eye blinking and IoT platform. In Proceedings of the 2020 IEEE 11th Latin American Symposium on Circuits Systems, San Jose, Costa Rica, 28 February 2020.

[25] Ren, P.; Ma, X.; Lai, W.; Zhang, M.; Liu, S.; Wang, Y.; Li, M.; Ma, D.; Dong, Y.; He, Y.; Xu, X. Comparison of the Use of Blink Rate and Blink Rate Variability for Mental State Recognition. IEEE Trans. Neural Syst. Rehabil. Eng. 2019, 27, 867–875.

[26] El Naqa, I.; Murphy, M.J. Machine learning in radiation oncology. In What is Machine Learning; Springer: Cham, Switzerland, 2005; pp. 3–11.

[27] Hu. S. Document image watermarking algorithm based on neighborhood pixel ratio. Proceedings. (ICASSP '05). In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, PA, USA, 23 March 2005.

[28] Suthaharan, S. Machine learning models and algorithms for big data classification. In Support Vector Machine; Springer: Boston, MA, USA, 2016; pp. 207–235.

[29] Albawi, S.; Mohammed, T.A.; Al-Zawi, S. Understanding of a convolutional neural network. In Proceedings of the International Conference on Engineering and Technology, Antalya, Turkey, 21–23 August 2017.

[30] Choi, K.; Cichocki, A. Control of a Wheelchair by Motor Imagery in Real Time. In International Conference on Intelligent Data Engineering and Automated Learning; Springer: Berlin/Heidelberg, Germany, 2008; pp. 330–337.

[31] Mahmoud, A.; Hamoud, M.; Ahmad, A.M.; Ahmad, A.S. Controlling a Wheelchair Using Human-Computer Interaction. Int. J. Sci. Res. 2018, 7, 681–686.

[32] Pande, V.V.; Ubale, N.S.; Masurkar, D.P.; Ingole, N.R.; Mane, P.P. Hand Gesture Based Wheelchair Movement Control for Disabled Person Using MEMS. J. Eng. Res. Appl. 2014, 4, 152–158.

[33] Chen, Y.L.; Chen, S.C.; Chen, W.L.; Lin, J.F. A head orientated wheelchair for people with disabilities. Disabil. Rehabil. 2003, 25, 249–253.

[34] Mandel, C.; Luth, T.; Navigating a smart wheelchair with a braincomputer interface interpreting steady-state visual evoked potentials. In Proceedings of the 2009 IEEE/RSJ International Conference, St. Louis, MO, USA, 10 October 2009.

[35] Lund, M.E.; Christiensen, H,V; Caltenco, H,A; Lontis, E.R.; Bentsen, B.; Andreasen Struijk, L.N.S. Inductive tongue control of powered wheelchairs. In Proceedings of the 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, Buenos Aires, Argentina, 1–4 September 2010.

[36] Singer, C.C.; Hartmann, B. See-Thru: Towards Minimally Obstructive Eye-Controlled Wheelchair Interfaces. In The 21st International ACM SIGACCESS Conference on Computers and Accessibility; ACM: New York, NY, USA, 2019.

Chapter 4. Experimental validation of intelligent recognition of eye movements in the application of autonomous vehicle driving

This paper presents a novel human-computer interaction system that has the potential to be used for autonomous or safety assisted driving in future vehicles. Convolutional neural networks were built and used for single shot detection and blink detection. The single shot detection method has been used to accomplish the detection of dynamic targets. The blink detection is performed by feeding multiple images of open and closed eyes into the network for deep learning, and based on the learned data models can detect the open or closed state of the subject's eyes. In addition, eye tracking technology is used to identify the direction of the driver's gaze. The human-computer interaction system is empirically validated in a super-compact electric vehicle, and it can accurately detect external dynamic targets, while the driver can control the vehicle by blinking and gaze direction.

4.1. Introduction

Autonomous driving is continuing to make significant progress in recent years with advances in various sensing and intelligent recognition technologies. A sensing system for autonomous vehicle navigation generally consists mainly of a combination of sensors, including active and passive sensors, i.e. cameras, radar and LIDAR [1-3]. LIDAR in particular acts as an active sensor, sensing the surroundings by emitting laser light. By processing the received laser echoes from the reflective surface, artificial intelligence techniques such as deep learning is applied to the precise measurement of distance for autonomous driving purposes. This is currently the dominant approach to autonomous driving. However, there are still some problems with the current LIDAR sensing approach. This is mainly reflected in the following aspects: (1) Due to the current speed of image processing and the current state of artificial intelligence technology, this kind of autonomous driving often requires autonomous driving on the basis of the known road environment; (2) The feasibility of LIDAR sensing on complex roads with frequent cross traffic and congestion has not yet been fully demonstrated. In particular, the safety of multiple vehicles emitting LIDAR waves in the event of mutual interference is still an important issue; (3) Any automatic control system with a supplementary safety operating system must be considered. Autonomous driving is no exception. It is therefore necessary to develop an autonomous driving approach that is different from LIDAR.

As an auxiliary safety system for autonomous driving, human-computer interaction (HCI) is considered to be one of the options. This is due to the potential to combine HCI with autonomous driving in the necessary environment, which even to replace LIDAR sensing for some of the time to implement autonomous driving to complement LIDAR sensing. It can also be used as an auxiliary safety system. For example, a normal camera senses the environment in front of the vehicle while it is moving, and this environment is almost identical to the one seen by the driver. Concurrently, research on human-computer interaction systems that do not rely on hands and feet is becoming widespread in order to improve the quality of life of people with disabilities [5]. Of particular interest are research efforts such as eye gaze and eye blink input that transmit information through the human eye [4-8]. Compared to other signals EEG and EMG signals etc., the signals transmitted by the eyes have better stability and real-time performance than other signals, as humans can communicate large amounts of information faster and more directly through the eyes [5-6]. In response, a human-computer interaction system that controls the movement of a computer or wheelchair by sensing human eye movements has been proposed [4-9]. The combination of these two sensory information forms the human-computer interaction system. Such a system has the potential to complement the laser sensing approach to autonomous driving.

Drawing on this idea, this paper combines human eye and blink movements with control to create a novel human-computer interaction system in a super-compact electric vehicle named as Toyota COMS electric vehicle (COMS EV). This system not only detects targets outside the cab, but also determines the blink or eye gaze direction of the driver inside the cab. In this paper, a convolutional neural networks (CNN) was constructed in this HCI system with reference to neural networks and deep learning theory in image processing [10-12], and used for single shot detection (SSD) [13] and blink detection [5]. The single shot detection method was used to accomplish the detection of dynamic targets. Blink detection is done by feeding multiple images of open and closed eyes into the network for deep learning and detecting the open or closed state of the human eye based on the learned data model. Similarly, we use eye-tracking technology to determine the direction of eye gaze. In this way, the driver can transmit information to the COMS EV by blinking and the direction of eye gaze, while the camera mounted on the COMS EV can detect targets such as people, cars and road signs outside the vehicle and feed back to the driver to give him an alert, while the vehicle itself can also perform early warning avoidance. A series of experiments were carried out to verify the effectiveness of this human-computer interaction system. The experimental results show that they are effective.

4.2. System design and composition

4.2.1. System Overview

The HCI control system in this paper includes, blink detection, eye gaze, and single shot detection. Figure 4-1 shows the HCI system for a vehicle used for assisted driving. The subject can send commands to the trip computer by blinking or eye gaze, which can control the movement of the vehicle, etc. Single-shot detection is used to determine the surrounding obstacles, feed the information to the subject and drive the vehicle to avoid the obstacles. In this study, this HCI system was installed on the COMS EV for experiments. The real-time image of the subject's eye movements were captured. The computer analyzed the captured eye images to obtain the information expressed by the subject and sent commands to the COMS EV based on the information expressed by the subject. The COMS EV received the commands from the computer and output the corresponding voltage signals to drive the vehicle movement. The single shot detection in the vehicle captures the conditions in the vehicle's driving route in real time, and when an obstacle is detected in front of it, it will provide timely warning feedback and drive the vehicle to avoid the obstacle.

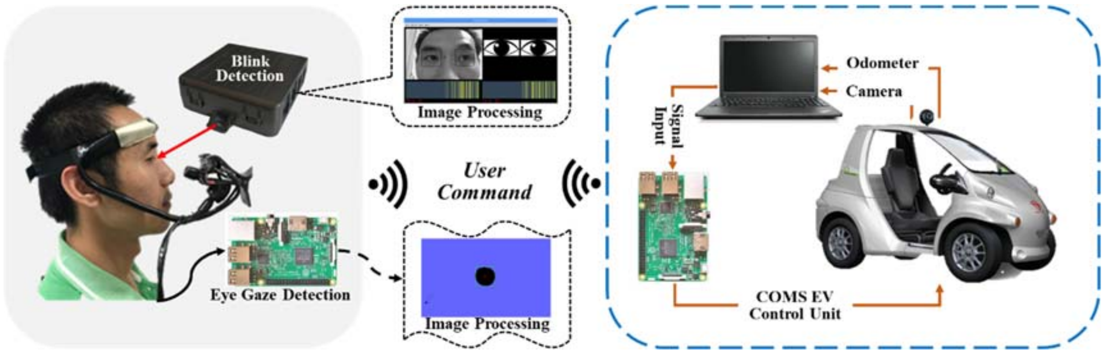


Figure 4-1. The HCI system of the vehicle for assisted driving.

4.2.2 Components of electric vehicle control system

The COMS EV was modified by adding gears to the steering lever of the EV and using an electric motor to turn the gears to control the direction of the car. Similarly, the brake pedal and accelerator pedal were equipped with motors to control the speed of the COMS EV by controlling the rotation of the motors. These motors are controlled by the ECU (Electronic Control Unit) developed in this study. Figure 4-2 shows the EV control system diagram.

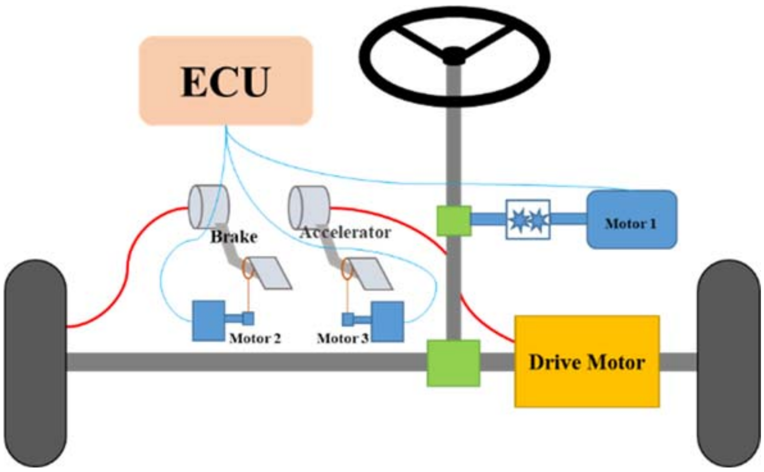


Figure 4-2. EV control system diagram

When the ECU receives a steering command, it controls the motor 1 rotation to drive the steering wheel through the gears, thus changing the direction of travel of the COMS EV. If the COMS EV needs to slow down or brake, the ECU drives the motor 2 to rotate and drive the brake pedal through the steel cable to slow down or brake, and drives the motor 3 to reset and release the accelerator pedal. In case the COMS EV needs to accelerate, the ECU drives the motor 3 to drive the accelerator pedal through the steel cable to accelerate, and drives the motor 2 to reset and release the brake pedal. In this way, the COMS EV driving can be controlled.

4.3 Image processing methods

4.3.1. Blink detection method

The blink detection in this study was done by processing and extracting features from images with open and closed eyes. For this purpose, the CNN was built to implement the feature extraction. The CNN is a kind of neural network mainly applied in the field of image recognition, which can effectively reduce the dimensionality of large data volumes into small data volume and can effectively retain the image features in accordance with the principle of image processing. Figure 4-3 shows the principle structure of the convolutional neural networks.

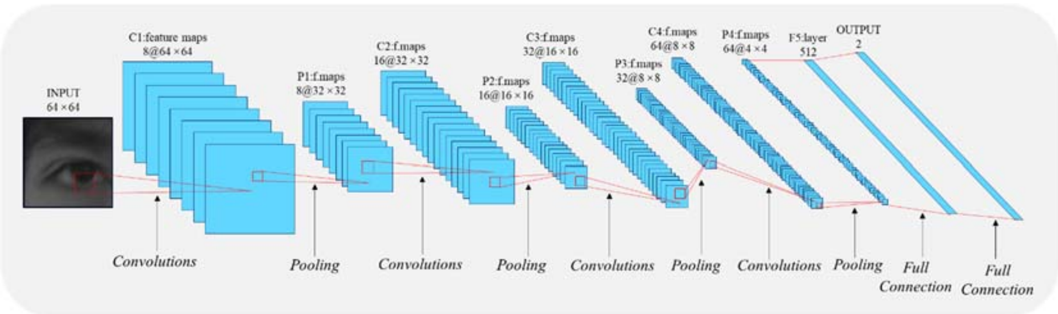


Figure 4-3. The principle structure of the convolutional neural networks.

The whole network consists of convolutional layers, pooling and fully connection layers. In this experiment, the well-known VGG-16 discriminative model is used as a reference. Since only two types of images (open-eye and closed-eye) are required for feature extraction and classification, the model was rescaled from the initial 16 layers to 10 layers in order to balance high efficiency and accuracy, and the experiments verified that the model being built is reliable.

In the convolution layer, a 3×3 convolution kernel is used for convolution operations and the ReLU function is used as the activation function. To keep the feature map size constant, one unit of padding is applied to the original feature map before each convolution operation. The 2×2 Max pooling is used for pooling with a stride of 2. The CNN ends with two fully connected layers. The important parameters of each CNN layer are shown in Figure 4-4.

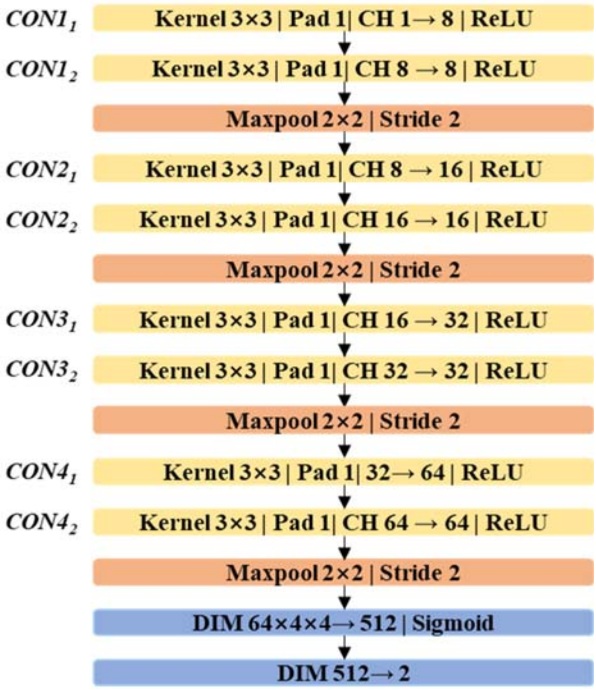


Figure 4-4. The relevant parameters in each layer of the CNN.

The padding is to increase the number of pixels on each edge. It is to not discard the original map information during the convolution process, and to maintain the feature map size consistent with the original map size, so that the inputs of the deeper layers remain large enough to have information. To achieve this purpose and to avoid mixing noise in the padding, the value of padding pixels setting by zero. Figure 4-5 shows the convolution process with padding.

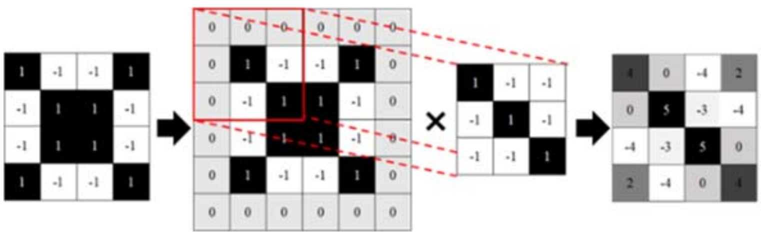


Figure 4-5. The convolution process with padding

A. Convolutional Layer

The convolutional layers are the core structure of the convolutional neural network. As shown in Figure 4-4, eight convolutional layers are used in the convolutional neural network we built. Each convolutional layer contains some meaningful features. A fixed size kernel is used for convolution to obtain the feature map, with the convolution process shown in equation (4-1).

$$x_{i,j}^{l} = f(\sum_m \sum_n w_{m,n}^{l} x_{i+m,j+n}^{l-1} + b_{m,n}^{l}) \tag{4-1}$$

Where $x_{i,j}^{l}$ is the element of the i-th row and j-th column of the l-layer feature map, f represents activation functions such as ReLU, Sigmoid, etc., $w_{m,n}^{l}$ denotes the m-th row and n-th column weights of layer l. The convolution kernel is determined by these trainable weights $w_{m,n}^{l}$. $b_{m,n}^{l}$ is the m-th row and n-th column bias of the feature map of the lth layer.

B. Pooling Layer

The schematic diagram of max pooling show in Figure 4-6. Pooling is the process of taking the maximum or average value over a range (e.g., 2x2). It is done by taking the maximum or average value of these four regions within the region of 2x2. After each completed stage of convolution, the pooling is performed. Max pooling has more robustness when ambient noise is present. Max pooling is used in this paper as it is more accurate. The output of each pooling is shown in equation (4-2).

$$x_{i,j}^{l} = f(w_{m,n}^{l} sam(x_{i+m,j+n}^{l-1}) + b_{n}^{l}) \tag{4-2}$$

Where *sam* ( ) denotes the sampling function, *w* denotes the weight coefficient, *b* is the bias of the output feature, and *f* is the activation function of the neuron (see in equation (4-2)).

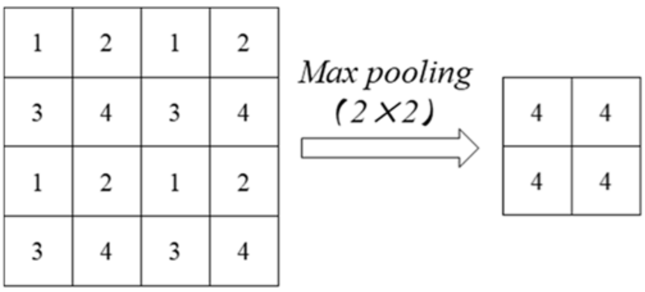


Figure 4-6. Schematic diagram of max pooling

C. Fully connection layer

The fully connected layer has each node connected to all the nodes in the previous layer and is used to combine all the features extracted earlier. Due to its fully connected nature, the fully connected layer will generate more parameters. The number of feature dimensions is reduced by convolutional layers to greatly reduce the computational effort. The output of each neuron is described in equation (4-3).

$$h_{W,b}(x) = f(W^T x + b \quad) \tag{4-3}$$

Where $h_{W,b}(x)$ is the output of neurons, x is the input neurons, W is the weight of connecting neurons, b is the bias, and f is the activation function of the neuron (see in equation (4-1) & (4-2)).

4.3.2. Eye gaze detection method

The principle of corneal reflection eye gaze tracking technique is shown in Figure 4-7, which uses contrast to locate the center of the pupil and uses infrared non-columnar light to produce a corneal reflection that covers the black area of the pupil by illuminating the eye with infrared light. In order to be able to use it in various light environments such as sunlight, a filter is added to the lens so that light other than infrared light can be filtered out.

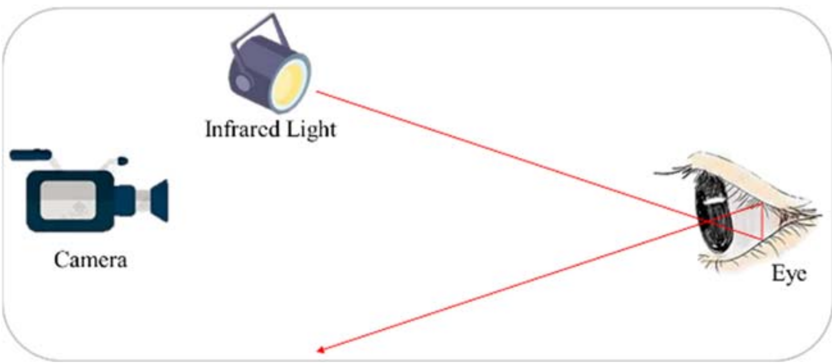


Figure 4-7. Diagram of corneal reflection eye tracking technology

The image binarization method is used to process captured eye images. Image binarization is a process that converts an image with shadows into two shades of black and white. In this process a threshold is preset and each pixel is replaced with white if its value is above the threshold and with black if it is below the threshold. The output of the binarisation is described in equation (4-4).

$$f'(x,y) = \begin{cases} 255 & f(x,y) \geq \theta \\ 0 & f(x,y) < \theta \end{cases}$$

(4-4)

Where the threshold is θ, the coordinate system is (x, y), the concentration value is $f(x, y)$, and the transformed concentration value is $f'(x, y)$.

The flow of the eye tracking technique is shown in Figure 4-8. After the eye image has been binarization processed, the pupil is extracted and the pupil center coordinates are calculated. The gaze direction can be calculated from the trajectory of the pupil center coordinates.

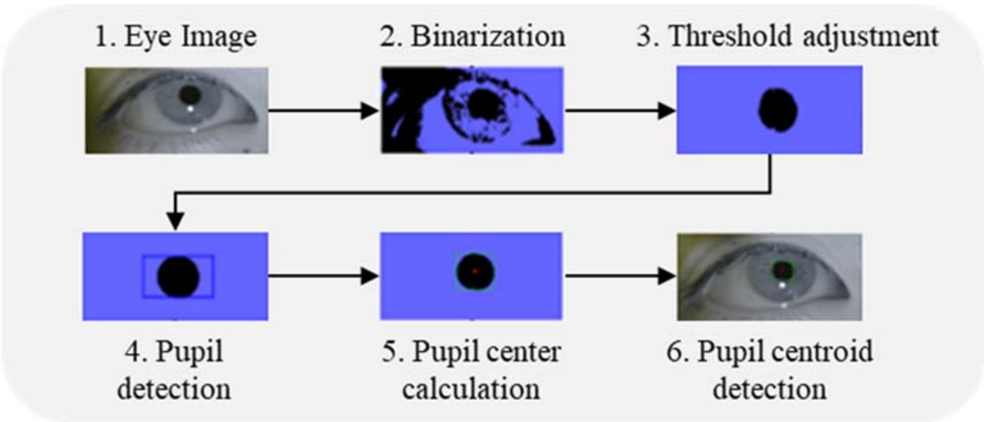


Figure 4-8. Flow chart of eyeball tracking technology

4.3.3. The eye movements and its GUI

The GUI of the blink detection system is shown in Figure 4-9. The GUI shows the camera image, the eye area, the blink waveform and the number of consecutive blinks by the user. In the top left area of the interface, the captured image is displayed. The cartoon eye image is displayed in a rectangular area. The state of this eye area is updated every 10 frames due to the processing speed. The two waveforms of the bottom area represent the open and closed states of the eyes, 1 for open eyes and 0 for closed eyes. The image in the upper right corner shows the opening or closing of the subject's eyes in the form of a moving picture. The symbols at the bottom indicate the number of consecutive blinks, with the symbol (*) increasing with each blink of the eye. We have introduced a face recognizer and an eye recognizer in the preset program, so that the capture frame in the screen always firmly captures both eyes and determines the direction of eye gaze according to the relative position of the pupils. When controlling the COMS EV driving, the number of separate blinks of the left and right eyes is compiled into different commands sent to the trip computer, which controls the car to do the relevant actions.

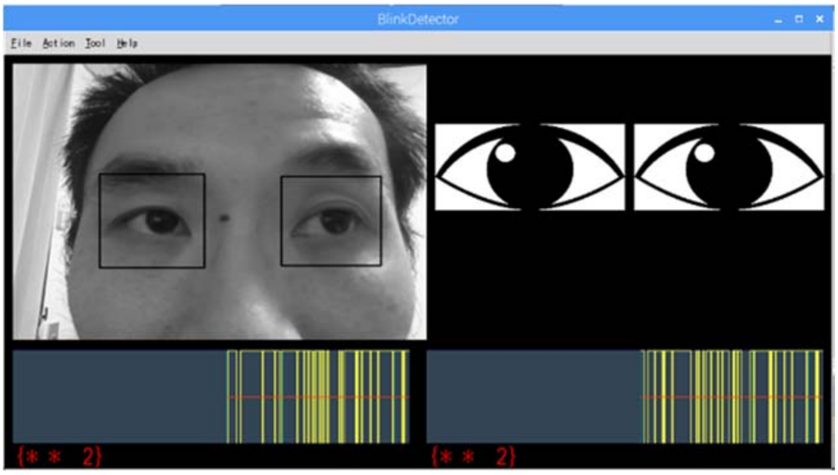


Figure 4-9. The GUI of the blink detection system

The GUI of the eye gaze detection system is shown in Figure 4-10. After the binarization process, the black pupil is extracted and a blue box is used to firmly lock the black pupil. When the pupil moves, the direction of eye gaze can be determined by calculating the unique area of the pupil. When interacting with the computer, the direction of eye gaze (i.e., the position of the pupil in the image) can be compiled into different commands and transmitted to the computer, which can then control the vehicle to do the corresponding actions according to the commands.

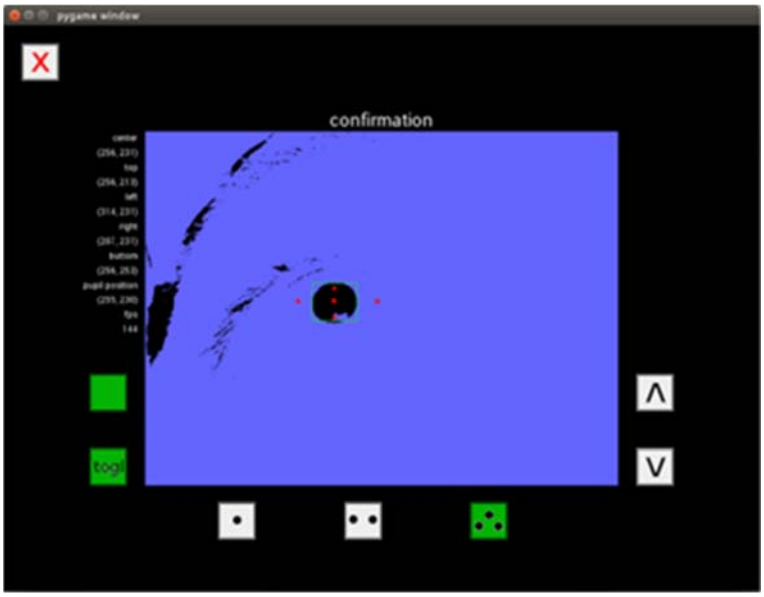


Figure 4-10. The GUI of the eye gaze detection system

4.4. Obstacle determination and motion control methods

4.4.1. The method of obstacle identification

The SSD is a dense sampling of uniformly different locations of the image, which can be sampled at different scales and aspect ratios. The CNN is then used to extract features for direct classification and regression, and the whole process takes one step, so its detection speed is very fast. In this paper, the VGG-16 is used as the base network model for SSD, and then a new convolutional layer is added to VGG-16 to obtain more feature maps for detection. Figure 4-11 shows the network structure of SSD.

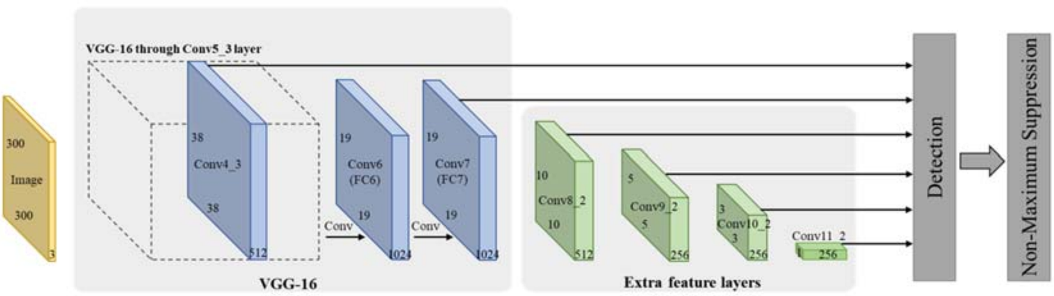


Figure 4-11. The network structure of SSD.

The core of SSD is to predict the box offset by a fixed set of category scores and default bounding boxes using a small convolutional filter applied to the feature map [13]. To achieve these, an image is first input, and the image is allowed to go through a convolutional neural network to extract features and generate a feature map with six layers of extracting features. Then a default box is generated on each pixel of the feature map. Finally, all the default boxes are collected and put into NMS (non-maximum suppression) for filtering, and the filtered default boxes are output. This allows the output of the ground truth (GT) box. Figure 4-12 shows the output principle of the ground truth box.

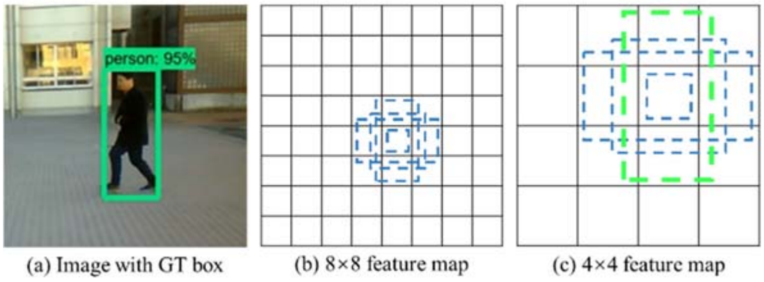


Figure 4-12. The output principle of the ground truth box

4.4.2. The motion control method

Figure 4-13 shows the control flow of the system. The subject can interact with the computer through the eyes. When the autonomous driving system encountered some impassable obstacles, the subject could avoid the obstacles by blinking and eye gaze to control the slow movement of the vehicle. The SSD system would provide real-time feedback on the surrounding road conditions and alert the subject in time. Once the vehicle is out of the way, the subject can use blinking and eye gaze to reselect the autopilot. The system also detects whether the subject is concentrated in real time through the subject's eyes while the vehicle is driving, and alerts the subject.

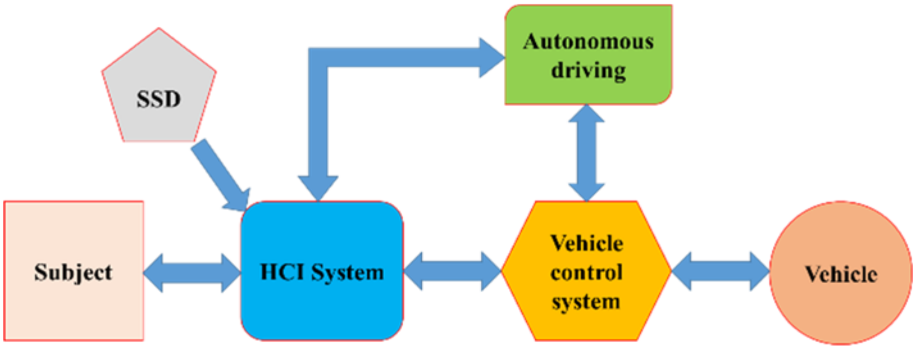


Figure 4-13. Control flow diagram of the system

The single shot detection about the obstacle distance is shown in Figure 4-14. The camera takes a picture with dimensions of length and width of 2*m* and 2*n* respectively, A (*x, y*) is the coordinates of the center point of the lower edge of the GT box of the detected object, and B (*m, n*) is the coordinates of the center point of the taken picture. By creating a similar triangle based on the principle that the size of the detected object in the image is proportional to the actual size of the detected object, the distance between the lens and the detected object can be calculated. In order to be accurate and not affected by the size of the detected object, the distance between the center point of the lower edge of the GT box of the detected object and the head of the vehicle is used as the judgment distance D. The distance D between the obstacle and the lens is calculated as in equation (4-5).

$$D = \frac{n \cdot cot\theta}{n - y} \cdot h - l \tag{4-5}$$

Where 2θ is the field of view of the lens, *h* is the vertical distance from the lens to the ground, and *l* is the horizontal distance from the lens to the front of the vehicle.

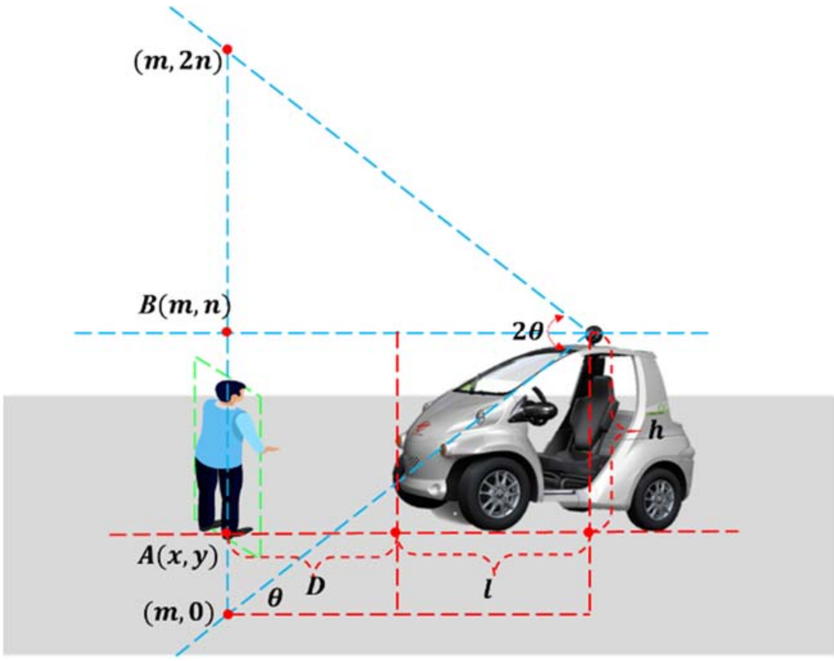


Figure 4-14. Distance judgment for the single shot detection

4.5. Experimental verification

The judged waveform of the blink detection experiment show in Figure 4-15. The subject performed 20 blinks in both eyes. Based on the data waveforms it can be seen that all 20 blinks in the left and right eyes were detected by the system.

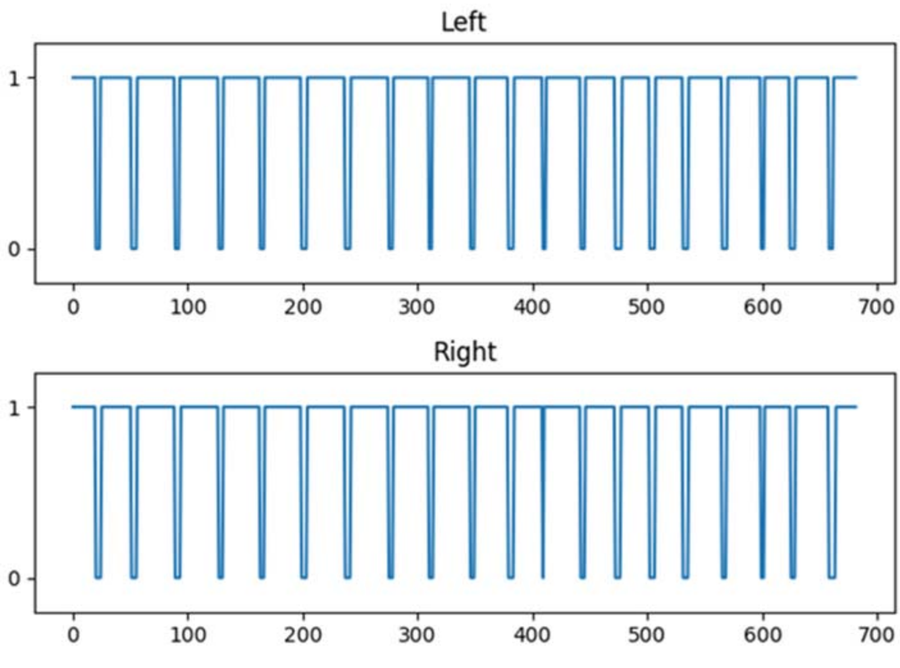


Figure 4-15. Judgment waveform for blink detection

In order to derive the accuracy of blink judgments, a sample of 10 subjects was taken. Each subject blinked 100 times continuously, for a total of 1000 consecutive blinks, and the system detected a total of 995 blinks with an accuracy of 99.5%. Table 1 shows the results of this experiment. Of these, eight subjects had all 100 blinks detected, one subject had three undetected blinks, and one subject had two undetected blinks. These two subjects' eyes showed more significant differences compared to the other subjects' eyes, and after recording the images of these two subjects' eyes into the training set and retraining them, the accuracy rate improved significantly and there were no more undetected blinks.

Table 1. Experiment with 10 subjects for a total of 1000 blink detections

Subject	Detection count	Undetected count	Misdetection count
Subject 1	100	0	0
Subject 2	100	0	0
Subject 3	100	0	0
Subject 4	100	0	0
Subject 5	97	3	0
Subject 6	100	0	0
Subject 7	100	0	0
Subject 8	100	0	0
Subject 9	98	2	0
Subject 10	100	0	0

The comparison of the eye images in different states show in Figure 4-16. Figure 4-16(a) shows the eye image without IR light irradiation, and Figure 4-16(b) shows the eye image with IR light irradiation. Comparing Figure 4-16(a) and Figure 4-16(b), it can be noticed that the contour of the pupil in Figure 4-16(b) is more clearly defined. Because corneal reflection is a way to detect pupil depth, it is possible to compare the color of the iris by separating the infrared light from the optical axis of the eye tracking camera. Figure 4-16(c) shows the eye image after the binarization process. To highlight the change of binarization, the background color is set to blue instead of white. After binarization, only two colors are left in the figure, and then the threshold is adjusted so that the pupil's outline becomes more visible in the image.

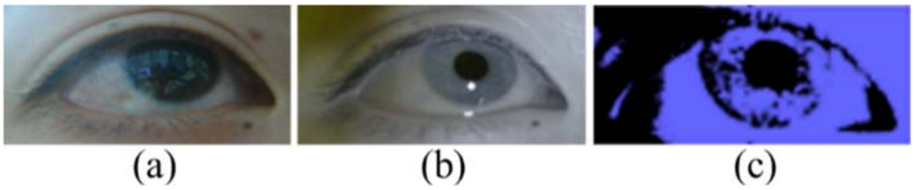


Figure 4-16. Comparison of eye images during the experiment

The SSD-based obstacle detection provides the driver with early warning information while the vehicle is in motion, alerting him to the presence of people or objects or traffic warning signs in the vicinity. Experiments have shown that distances of up to 40 meters can be detected accurately and quickly. Figure 4-17 shows the dynamic detection of experimental images.

Figure 4-17. Dynamic detection of experimental images

The experiment to control the motion of the COMS EV show in Figure 4-18. Firstly, the COMS electric vehicle was allowed to drive automatically at a speed of 5 km/h with the subject standing in front of the left side of the vehicle, and the vehicle automatically turned to the right as it passed to go around the subject (see No. 1-3 in Figure 4-18). Then, the subject was asked to quickly stand in front of the right side of the vehicle, which automatically turned left to avoid the subject as it passed (see No. 4-6 in Figure 4-18). Finally, the subject quickly stood less than 2 meters from the front of the vehicle, and the vehicle took emergency braking in this case because it could not automatically turn to avoid the subject (see No. 7-9 in Figure 4-18). After five iterations of the experiment, the COMS EV was able to successfully complete the task each time.

Figure 4-18. The experiment of controlling the movement of the COMS EV

In the next step, another subject was allowed to control the vehicle by eye gaze and blinking and drove at 5 km/h on the previous path, easily avoiding the obstacles and reaching the end point with success (see NO. 10-12 in Figure 4-18). In total, three experiments were conducted, and each time the subjects were able to control the COMS EV to successfully avoid the obstacles and complete each experiment successfully.

4.6. Conclusions

In this paper, the application of eye movements intelligence recognition in autonomous vehicle driving is investigated. A system incorporating obstacle recognition, eye movement recognition assisted driving is developed. Eye movement recognition is divided into blink detection and eye gazes direction detection. Blink detection uses convolutional neural networks to extract eye opening and eye closing features to determine blinks. Eye gaze direction uses binarisation of the eye image to locate the center of the pupil and thus determine the direction of eye gaze. Obstacle detection uses SSD to discern the type of obstacle and to give an early warning.

The system was transferred to the Toyota COMS EV for experiments, where subjects successfully avoided obstacles by controlling the vehicle's movement through their eyes on a set course. The SSD based obstacle detection system also worked on the experiment, providing feedback to the subjects about the conditions ahead of them on their route and enabling them to avoid obstacles. At the same time, the system was able to determine whether the subject was paying attention while driving based on the subject's blinks and eye gazes, and reminded the subject to pay attention. The experimental results show that the driving aid system is effective.

The safety and reliability of the detection will be further improved in the future, and the training time will be shortened to allow more subjects to perform the experiments and further improve the learning efficiency. This will enable the intelligent car driving assistance system with human-computer interaction to be widely used so that it can better serve people with disabilities.

References

[1] Veli, I and Charles, T.: “High Definition 3D Map Creation Using GNSS/IMU/LiDAR Sensor Integration to Support Autonomous Vehicle Navigation”, Sensors, Vol.20, No.3, pp.899., 2020.

[2] Li, Y and Ibanez-Guzman, J.: “Lidar for autonomous driving: The principles, challenges, and trends for automotive lidar and perception systems”, IEEE Signal Processing Magazine, Vol.37, No.4, pp.50-61., 2020.

[3] Gao, H., Cheng, B., Wang, J., Li, K., Zhao, J and Li, D.: “Lidar for autonomous driving: The principles, challenges, and trends for automotive lidar and perception systems”, IEEE Transactions on Industrial Informatics, Vol.14, No.9, pp.4224-4231., 2018.

[4] Utaminingrum, F., Purwanto, A., Masruri, M., Ogata, K., Somawirata, I.: “Eye Movement and Blink Detection for Selecting Menu On-Screen Display Using Probability Analysis Based

on Facial Landmark", in International Journal of Innovative Computing, Information and Control, vol.17, no.4, pp.1287-1303., 2021.

[5] Luo, W., Cao, J., Ishikawa, K., Ju, D.: "A Human-Computer Control System Based on Intelligent Recognition of Eye Movements and Its Application in Wheelchair Driving", Multimodal Technologies and Interaction, vol.5, no.9, pp.50., 2021.

[6] Cojocaru, D., Manta, L., Vladu, I., Dragomir, A., Mariniuc, A.: "Using an eye gaze new combined approach to control a wheelchair movement", 23rd International Conference on System Theory, Control and Computing (ICSTCC), pp.626-631., 2019.

[7] Hutchinson, T., White, K., Martin, W., Reichert, K., Frey, L.: "Human-computer interaction using eye-gaze input", IEEE Transactions on systems, man, and cybernetics, Vol.19, No.6, pp.1527-1534., 1989.

[8] Laddi, A and Prakash, N.: "Eye gaze tracking based directional control interface for interactive applications", Multimedia Tools and Applications, Vol.78, No.22, pp.31215-31230., 2019.

[9] Fletcher, L and Zelinsky, A.: "Driver inattention detection based on eye gaze—Road event correlation", The international journal of robotics research, Vol.28, No.6, pp.774-801., 2009.

[10] Lawrence, S., Giles, C., Tsoi, A., Back, A.: "Face recognition: A convolutional neural-network approach", IEEE transactions on neural networks, Vol.8, No.1, pp.98-113., 1997.

[11] Kumar, A and Srivastava, S.: "Object Detection System Based on Convolution Neural Networks Using Single Shot Multi-Box Detector", Procedia Computer Science, Vol.171, pp.2610-2617., 2020.

[12] Albawi, S., Mohammed, T., Al-Zawi, S.: "Understanding of a convolutional neural network", International Conference on Engineering and Technology (ICET), pp.1-6., 2017.

[13] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C and Berg, A.: "Ssd: Single shot multibox detector", European conference on computer vision, pp.21-37., 2016.

Chapter 5. Conclusions

This paper investigates the application of combining the extraction of eye movement information to control relevant mechanical movements and eye movement intelligence recognition in autonomous vehicle driving. A human-computer interaction system that combines obstacle recognition and eye movement recognition is developed. Eye movement recognition is divided into blink detection and eye gaze direction detection. Blink detection uses convolutional neural networks to extract eye opening and eye closing features to determine blink. Eye gaze direction uses binarization of the eye image to locate the center of the pupil and thus determine the direction of eye gaze. Obstacle detection uses SSD to identify the type of obstacle and issue early warnings. The system was ported to a power wheelchair and a Toyota COMS cart for experimental validation.

For the experimental validation part of controlling the wheelchair using this HCI system, we first modified the wheelchair hardware to develop a hardware drive system for the electric wheelchair. This allowed the wheelchair to receive commands from the subject through eye movements. Blinking and the direction of gaze of the eyes were used to drive the wheelchair movement. The wheelchair was controlled and followed a set route, successfully avoiding obstacles along the way and reaching the end point, and was tested repeatedly and successfully. The experiments proved that the method of obtaining information expressed by the human eye through image parsing and machine learning is effective. The conclusions are summarized as follows.

(1) An eye gaze detection system based on image recognition and eye tracking methods was developed. The system can use eye movements to control vehicles such as wheelchairs.

(2) A blink detection system based on an image recognition method has been developed. It switches vehicle gears and emergency braking by the subject's blink.

(3) A PIC-based hardware circuit system was designed and developed to drive the wheelchair movement after receiving the command.

(4) A wheelchair human-computer interaction system based on intelligent eye-movement recognition was developed. It can be used by disabled people and ALS patients to travel.

Similarly, the HCI system system was equipped as an assisted safety driving system on a Toyota COMS minivan, and the subject successfully avoided obstacles by controlling the vehicle's movement on a set route with his eyes. The SSD-based obstacle detection system also worked in the experiment, giving subjects feedback on what was ahead of their route, enabling them to avoid obstacles. At the same time, the system was able to determine whether the subjects were paying attention while driving based on their blinks and eye gaze, and alerted the subjects to pay attention. The experimental results show that the driving assistance system is effective. The conclusions are summarized as follows.

(1) A target identification and target distance detection system was developed. It is used for obstacle recognition and detection in smart driving vehicles to improve the safety of driving.

(2) We have developed an ECU for COMS EV, which can drive the EV after receiving commands from outside via TCP/IP.

(3) An intelligent eye-movement recognition-based assisted driving system has been developed. The system detects the driver's fatigue level as well as attention, and uses eye movements to control the vehicle movement when there are some roads where autonomous driving does not work.

In the future, the safety and reliability of the detection will be further improved, and the training time will be shortened to allow more subjects to conduct the experiment and further improve the learning efficiency. This will lead to the widespread application of new vehicle human-computer interaction systems to better serve people with physical impairments.

Chapter 6. Future Studies

6.1. Intelligent wheelchairs combining multimodal HCI and SLAM

Combining this wheelchair HCI system with camera SLAM technology to develop a more intelligent wheelchair system. It can be used to plan paths through HCI or control vehicles to unfamiliar places and build SLAM maps. Once a SLAM map of that path has been created, the path can be selected by the HCI system and the smart wheelchair will then drive itself to the planned destination.

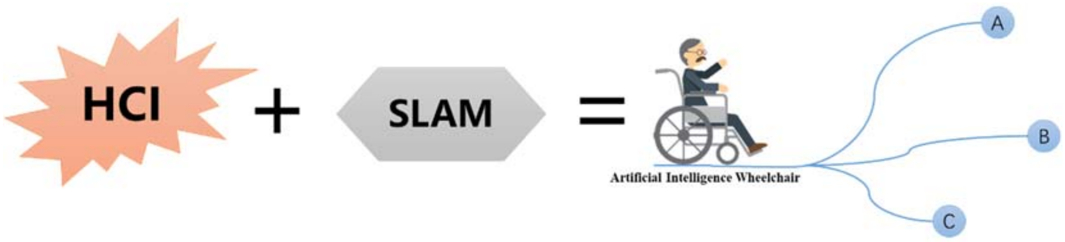


Figure 6-1. An overview of intelligent wheelchairs combining multimodal HCI and SLAM technologies

6.2. The future of intelligent vehicles

Application of the HCI system of this EV in combination with LIDAR SLAM and V2X technology to autonomous vehicles. This allows for the exchange of information about people and vehicles, and vehicles to the outside world.

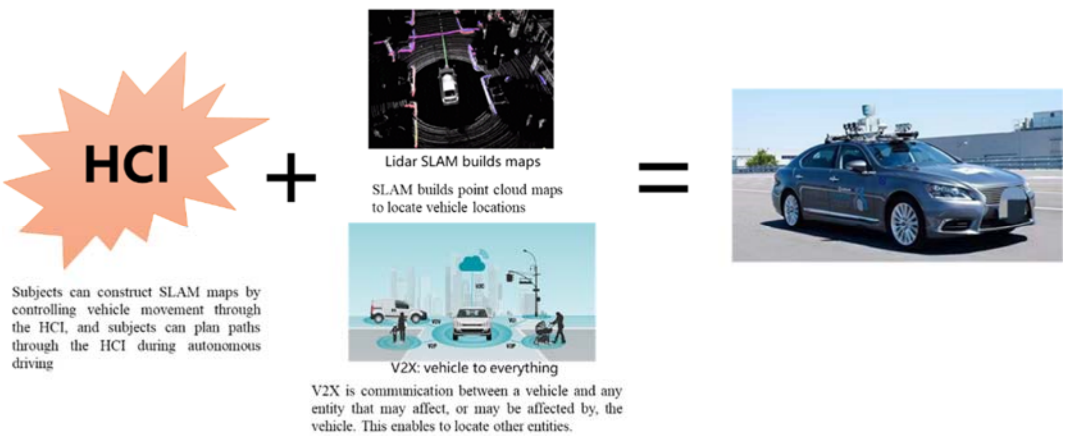


Figure 6-2. An overview of the future of intelligent vehicles

Related publications

Academic Journal:

Wenping Luo, Jianting Cao, Kousuke Ishikawa, Dongying Ju: A human-computer control system based on intelligent recognition of eye movements and its application in wheelchair driving, *Multimodal Technologies and Interaction*, Vol 5, No 9, pp 50-64, August 2021.

Wenping Luo, Jianting Cao, Kousuke Ishikawa, Dongying Ju: Experimental validation of intelligent recognition of eye movements in the application of autonomous vehicle driving, *International Journal of Biomedical Soft Computing and Human Sciences*, Vol 26, No 2, pp 63-72, December 2021.

International Conference

Wenping Luo, Zhongxu Tai, Dongying Ju, Jianting Cao: Development of a new type of multifunctional vanadium redox flow energy storage system and application proposal in distributed smart grid, *4th International Conference on New Energy and Future Energy Systems (NEFES 2019)*, July 21-24, 2019, Macao.

Wenping Luo, Zhongxu Tai, Dongying Ju, Jianting Cao: A grid-connected photovoltaic power generation and energy storage system based on deep reinforcement learning, *Applied Energy Symposium and Forum 2020: Low carbon cities and urban energy systems (CUE2020)*, October 10-17, 2020, Tokyo.

Acknowledgement:

I would like to express my sincere gratitude to my supervisors, Prof. Dong-Ying Ju and Prof. Jianting Cao, for accepting me as a student and giving me the opportunity to do research at Saitama Institute of Technology. I would like to sincerely thank them for their support, encouragement and guidance during my study. This thesis would not have been possible without their tireless efforts to revise and embellish my drafts. They spent most of their time reading my manuscript and suggesting further revisions. In addition, Professor Dong-Ying Ju has given me a lot of support and generous help in securing research funding, seeking corporate assistance, and in my life, for which I would like to express my sincere gratitude in particular.

I also would like to express my sincere gratitude to Prof. Takaharu Yamazaki, Prof. Tomomi Hashimoto and Prof. Wataru Oyama their guidance and help with this research.

In my research works, I would also like to sincerely thank Associate Professor Taiyo Maeda, Mr. Hiroki Ishizaki, and Dr. Yukisuke Ishikawa, who have helped me a lot in conducting my experimental research.

In addition, the research group members of the Prof. Ju's laboratory, M.S. Jianbin Lu, M.S. Yu Qiao, B.S. Toshio Arai, and B.S. Tatsuo Yamagata, have also assisted me in some experimental work, and I would like to express my special thanks to them again. I also thank all the members of Prof. Ju's laboratory and Prof. Cao's laboratory for their pleasant help and cooperation over the years.