ISSN: 2414-1895 | DOI: 10.6919/ICJE.202503_11(3).0020

The Working Area is Multiintelligent Security Monitoring and Space Electronic Fence

Jiahui Zhang[1], Wenxi Zhang[1], Yvshuo You[2], Jiang Ma[1, *]

[1] School of Science, North China University of Science and Technology, Tangshan 063000, China

[2] School of Mechanical and Electrical Engineering, Central South University, Changsha 410000, China

Abstract

In the mining and ore processing industry, the operation of truck unloading needs man-machine cooperation, which has huge safety and efficiency problems. The purpose of this study is to create an intelligent isolated driving system and a security system for human space reconstruction based on dynamic multi-perspective. For the work area, multi-camera human body recognition is carried out, 3D reconstruction is carried out according to the camera position, and the specific position of the 3D model is transmitted back, and the personnel entering the dangerous area is alerted and intelligent devices are assisted in positioning and operation. Thus improve safety efficiency and ensure the safety of plant workers.

Keywords

Efficiency; Intelligent; Safety.

1. Introduction

This research focuses on the development and optimization of two key safety monitoring systems: intelligent isolated driving system and dynamic multi-perspective human space reconstruction security system.

The intelligent isolated driving system realizes real-time monitoring and tracking of personnel in the working area through high-precision sensors and real-time image processing technology, ensuring that the driving and personnel maintain a safe distance and avoid human-machine conflict. The system uses advanced vision technology, including high-resolution cameras and sensor networks, combined with a central processing unit for real-time data processing and feedback, providing an intuitive visual operation interface.

The human space reconstruction security system based on dynamic multi-perspective uses multiple cameras to capture dynamic data from different angles, combines computer vision technology and deep learning algorithm to accurately reconstruct the position of the human body in three-dimensional space, and enhances the safety of factory workers. Through triangulation method and deep learning algorithm, the system realizes the recognition of key feature points of human body, the reconstruction of three-dimensional model, and the estimation of human posture. In addition, the study will explore the application of the latest version of YOLO's YOLOv8 algorithm in object detection, which performs well in real-time detection tasks with its speed and accuracy, by training YOLOv8 models to improve their performance in specific scenarios, and integrating them into our security systems. The ultimate goal of the research is to improve the safety of the working environment and increase production efficiency through technological innovation.

Volume 11 Issue 3, 2025
DOI: 10.6919/ICJE.202503_11(3).0020

2. Organization of the Text

2.1 Research Technology

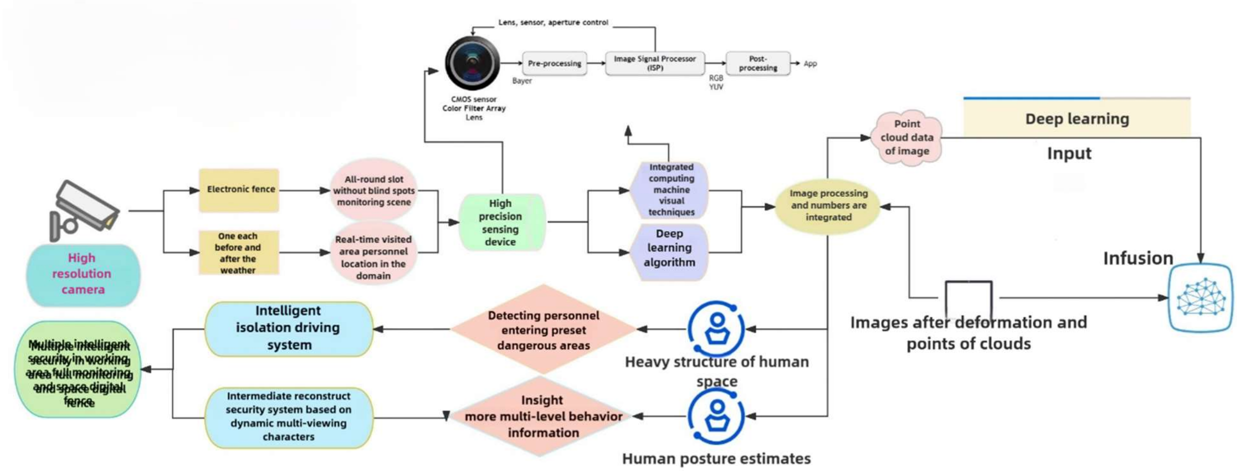

Fig. 1 Research technology route

2.1.1 YOLO Model

YOLO is a popular object detection algorithm whose core idea is to transform the object detection task into a single inference problem. Unlike traditional object detection methods, YOLO algorithm simultaneously predicts the target location and category in the image through a unified neural network. Real-time object detection has become an important part of many applications, spanning autonomous vehicles, robotics, video surveillance and augmented reality. Among the various object detection algorithms, the YOLO (You Only Look Once) framework stands out for its remarkable balance of speed and accuracy to quickly and reliably identify objects in images. Since its inception, the YOLO series has gone through many iterations, each time addressing limitations and improving performance on previous versions. Here, we are using YOLOv8 in YOLO, the eighth version of the YOLO model, which was proposed by Ultralytics researchers and which makes some improvements over YOLOv5: Stronger feature extractors YOLOv8s, YOLOv8m, YOLOv8l and YOLOv8x were used to enhance feature expression by using residual structure and attention mechanism; Used a variety of data enhancement methods, such as Mosaic, CutMix, MixUp, etc., to improve the generalization ability of the model; Various optimization methods, such as SGD and CosineAnnealing, have been used to accelerate the convergence of the model; Various detection techniques, such as IoU loss and NMS, are used to improve the accuracy and speed of the model.

YOLOv8 has the advantages of high precision, fast speed, and can detect objects of multiple scales and multi-labels. Moreover, the model is small and requires less computing resources, and the structure of the model is relatively simple, easy to use and deploy. However, the disadvantage is that the detection of small objects is still not good enough, and the performance and stability of the model need to be verified. Compared with YOLOv5,YOlOv8 eliminates the object loss in the loss function and uses classification loss instead, which not only improves the utilization rate of the model weight, but also makes the label better guide the model to learn the classification ability through the strong correlation between existence state and classification state. YOLOv8 uses BCE (binary cross entropy) as a classification loss, each category determines "is this class", and outputs the confidence level.

$$L = \frac{1}{N}\sum_{i} L_i = \frac{1}{N}\sum_{i} -\left[y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)\right] \tag{1}$$

In terms of regression loss, although the use of CIOU in v8 is the same as that in v5, the introduction of Anchor-Free center-based methods(based on the Center point) can reduce the risk of loss. The

ISSN: 2414-1895 DOI: 10.6919/ICJE.202503_11(3).0020

model changes from the output "offest of anchor frame size" to "predict the distance of the left, top, right, and bottom borders of the target frame from the target center point (ltrb = left, top, right, bottom)". In order to match Anchor-Free and improve generalization, DFL loss has been added in v8. In the form of cross entropy, DFL optimizes the probability of the two positions closest to the label y, so that the network can focus on the target location and the distribution of adjacent regions more quickly. This loss of position around the learning label can enhance the generalization of the model in complex situations, such as blocking and moving objects.

$$\mathrm{DFL}(S_i, S_{i+1}) = -\left((y_{i+1} - y)\log(S_i) + (y - y_i)\log(S_{i+1})\right) \tag{2}$$

2.1.2 Multi-view Image 3D Reconstruction Method

Multi-camera data acquisition and real-time video stream processing are the key technologies to achieve accurate monitoring and security[1]. First, the system uses multiple high-resolution cameras to simultaneously capture the scene from different angles, ensuring a full range of viewing angles with no blind spots. The following figure shows the preliminary camera layout:

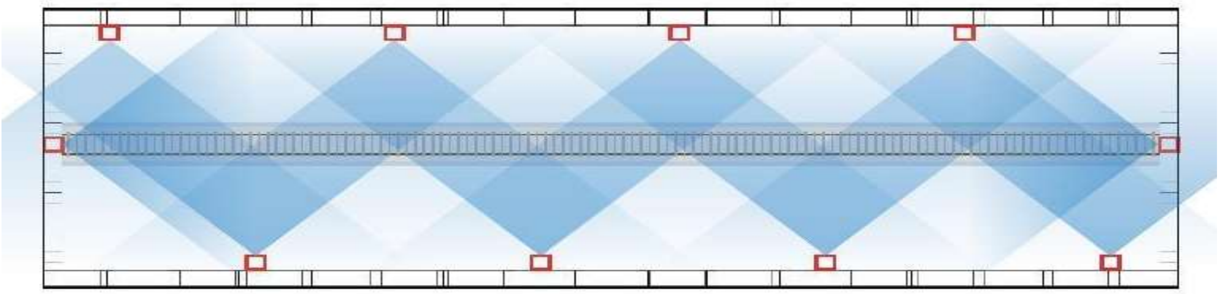

Fig. 2 All-round high-resolution camera layout

These cameras are carefully positioned to maximize scene coverage and optimize viewing angles, especially for activity capture in critical operational areas, enabling efficient and comprehensive monitoring. In terms of video data processing, the system ensures that the video data stream collected from each camera can be perfectly synchronized. This synchronous operation provides an accurate timing basis for image analysis and data fusion, ensuring the consistency and accuracy of information processing.

The basic principle of 3D reconstruction is to restore the 3D structure of an object through multi-view images or depth information[错误!未找到引用源。]. Common algorithms include stereovision, Structure from Motion (SfM), and multi-view stereogeometry. The core idea of these algorithms is to use the information difference between multi-view images to extract depth information, and then restore the three-dimensional structure of the object. Fig. 3 shows a simple imaging model of the camera[3]. A point P in Fig. 3, represented as Pw in the world coordinate system Ow-XwYwZw, is transformed into the camera coordinate system Oc-XcYcZc after rotation and translation operations, represented as Pc. After the projection, the transformation to the image plane coordinate system o 15-XY is represented as p fifteen (x, y), and the conversion from three-dimensional space to the image plane is completed. Finally, it is mapped to the pixel coordinate system through affine transformation in the same plane, and the point p (u, v) is the mapping point of P in the pixel plane.

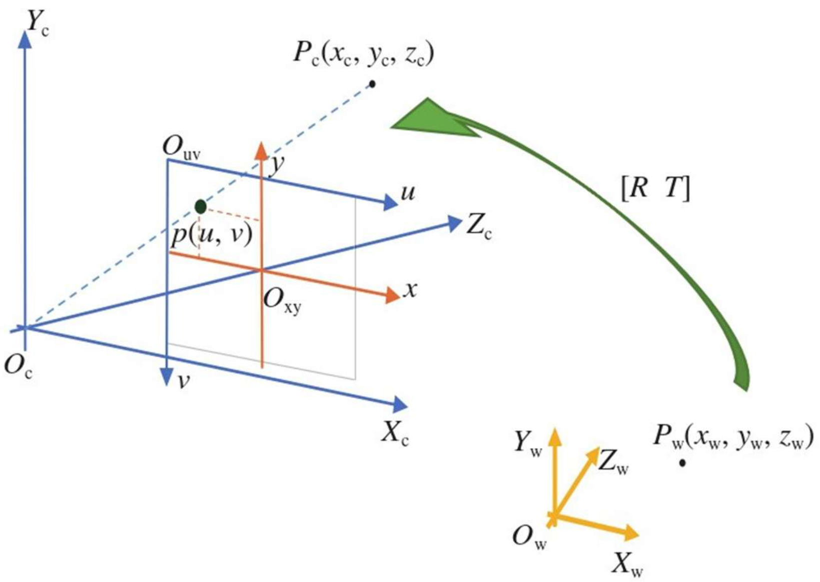

Fig. 3 Camera imaging model

Rotation and translation transformations from the world coordinate system to the camera coordinate system are:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & R_{13} & T_x \\ R_{21} & R_{22} & R_{23} & T_y \\ R_{31} & R_{32} & R_{33} & T_z \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{3}$$

Where [R T] is a 3×4 external parameter matrix composed of rotation matrix and translation vector. The transformation from camera coordinate system to pixel coordinate system is:

$$z_c \begin{bmatrix} u \\ v \\ l \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_x & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} \tag{4}$$

To wit:

$$\begin{bmatrix} u \\ v \\ l \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_x & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c / z_c \\ y_c / z_c \\ 1 \end{bmatrix} \tag{5}$$

Where: f x and f y represent the focal length of the camera on the X and Y axes; (u0,v0) is the primary point coordinate. If, represents the coordinates of point P in some normalized camera coordinate system. The collinear equation for a monocular camera is:

ISSN: 2414-1895 DOI: 10.6919/ICJE.202503_11(3).0020

$$\begin{cases} x_c^{'} = \dfrac{x_c}{z_c} = \dfrac{R_{11}x_w + R_{12}y_w + R_{13}x_w + T_x}{R_{31}x_w + R_{32}y_w + R_{33}x_w + T_z} \\ y_c^{'} = \dfrac{y_c}{z_c} = \dfrac{R_{21}x_w + R_{22}y_w + R_{23}x_w + T_y}{R_{31}x_w + R_{32}y_w + R_{33}z_w + T_z} \end{cases} \tag{6}$$

The position information of target points in different images is needed to produce stereo vision based on 3D reconstruction of images. The corresponding points generated by binocular cameras are shown in the fig. 4.

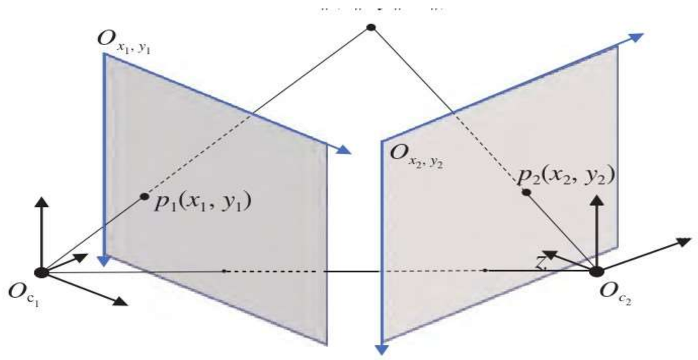

Fig. 4 Schematic diagram of corresponding points generated by binocular camera

Point P in Fig. 4: Point p 1 and p 2 of two images are generated by two different cameras, and four collinear equations are generated according to equation (6) :

$$\begin{cases} x_{c1}^{'} = \dfrac{R_{11}^{(1)}x_w + R_{12}^{(1)}y_w + R_{13}^{(1)}z_w + T_x^{(1)}}{R_{31}^{(1)}x_w + R_{32}^{(1)}y_w + R_{33}^{(1)}z_w + T_z^{(1)}} \\ y_{c1}^{'} = \dfrac{R_{21}^{(1)}x_w + R_{22}^{(1)}y_w + R_{23}^{(1)}z_w + T_y^{(1)}}{R_{31}^{(1)}x_w + R_{32}^{(1)}y_w + R_{33}^{(1)}z_w + T_z^{(1)}} \\ x_{c2}^{'} = \dfrac{R_{11}^{(2)}x_w + R_{12}^{(2)}y_w + R_{13}^{(2)}z_w + T_x^{(2)}}{R_{31}^{(2)}x_w + R_{32}^{(2)}y_w + R_{33}^{(2)}z_w + T_z^{(2)}} \\ y_{c2}^{'} = \dfrac{R_{21}^{(2)}x_w + R_{22}^{(2)}y_w + R_{23}^{(2)}z_w + T_y^{(2)}}{R_{31}^{(2)}x_w + R_{32}^{(2)}y_w + R_{33}^{(2)}z_w + T_z^{(2)}} \end{cases} \tag{7}$$

Where, R (1), T(1), R (2), T(2) respectively represent the rotation and translation transformation of the first and second camera coordinate system relative to the world coordinate system; Open parenthesis subscript base, x... Raise it to the ', end the baseline, subscript c 1,y subscript c 1... Power 'device control 4, close parenthesis and left parenthesis subscript baseline, x... Raise it to the ', end the baseline, subscript c 2,y subscript c 2... Raised to the power 'device control 4, the closing parenthesis indicates the coordinates of the point in the first and second normalized Ax = 0 camera coordinate system.

ISSN: 2414-1895
DOI: 10.6919/ICJE.202503_11(3).0020

$$\left|\begin{matrix} x_{c1}{}' R_{31}^{(1)} - R_{11}^{(1)} & x_{c1}{}' R_{32}^{(1)} - R_{12}^{(1)} & x_{c1}{}' R_{33}^{(1)} - R_{13}^{(1)} & x_{c1}{}' T_{z}^{(1)} - T_{x}^{(1)} \\ y_{c1}{}' R_{31}^{(1)} - R_{21}^{(1)} & y_{c1}{}' R_{32}^{(1)} - R_{22}^{(1)} & y_{c1}{}' R_{33}^{(1)} - R_{23}^{(1)} & y_{c1}{}' T_{z}^{(1)} - T_{y}^{(1)} \\ x_{c2}{}' R_{31}^{(2)} - R_{11}^{(2)} & x_{c2}{}' R_{32}^{(2)} - R_{12}^{(2)} & x_{c1}{}' R_{33}^{(2)} - R_{13}^{(2)} & x_{c1}{}' T_{z}^{(2)} - T_{x}^{(2)} \\ y_{c2}{}' R_{31}^{(2)} - R_{21}^{(2)} & y_{c2}{}' R_{32}^{(2)} - R_{22}^{(2)} & y_{c2}{}' R_{33}^{(2)} - R_{23}^{(2)} & x_{c2}{}' T_{z}^{(2)} - T_{y}^{(2)} \end{matrix}\right| \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{8}$$

$$A = \begin{bmatrix} x_{c1}{}' R_{31}^{(1)} - R_{11}^{(1)} & x_{c1}{}' R_{32}^{(1)} - R_{12}^{(1)} & x_{c1}{}' R_{33}^{(1)} - R_{13}^{(1)} & x_{c1}{}' T_{z}^{(1)} - T_{x}^{(1)} \\ y_{c1}{}' R_{31}^{(1)} - R_{21}^{(1)} & y_{c1}{}' R_{32}^{(1)} - R_{22}^{(1)} & y_{c1}{}' R_{33}^{(1)} - R_{23}^{(1)} & y_{c1}{}' T_{z}^{(1)} - T_{y}^{(1)} \\ x_{c2}{}' R_{31}^{(2)} - R_{11}^{(2)} & x_{c2}{}' R_{32}^{(2)} - R_{12}^{(2)} & x_{c2}{}' R_{33}^{(2)} - R_{13}^{(2)} & x_{c2}{}' T_{z}^{(2)} - T_{x}^{(2)} \\ y_{c2}{}' R_{31}^{(2)} - R_{22}^{(2)} & y_{c2}{}' R_{32}^{(2)} - R_{22}^{(2)} & y_{c2}{}' R_{33}^{(2)} - R_{23}^{(2)} & y_{c2}{}' T_{z}^{(2)} - T_{y}^{(2)} \end{bmatrix} \tag{9}$$

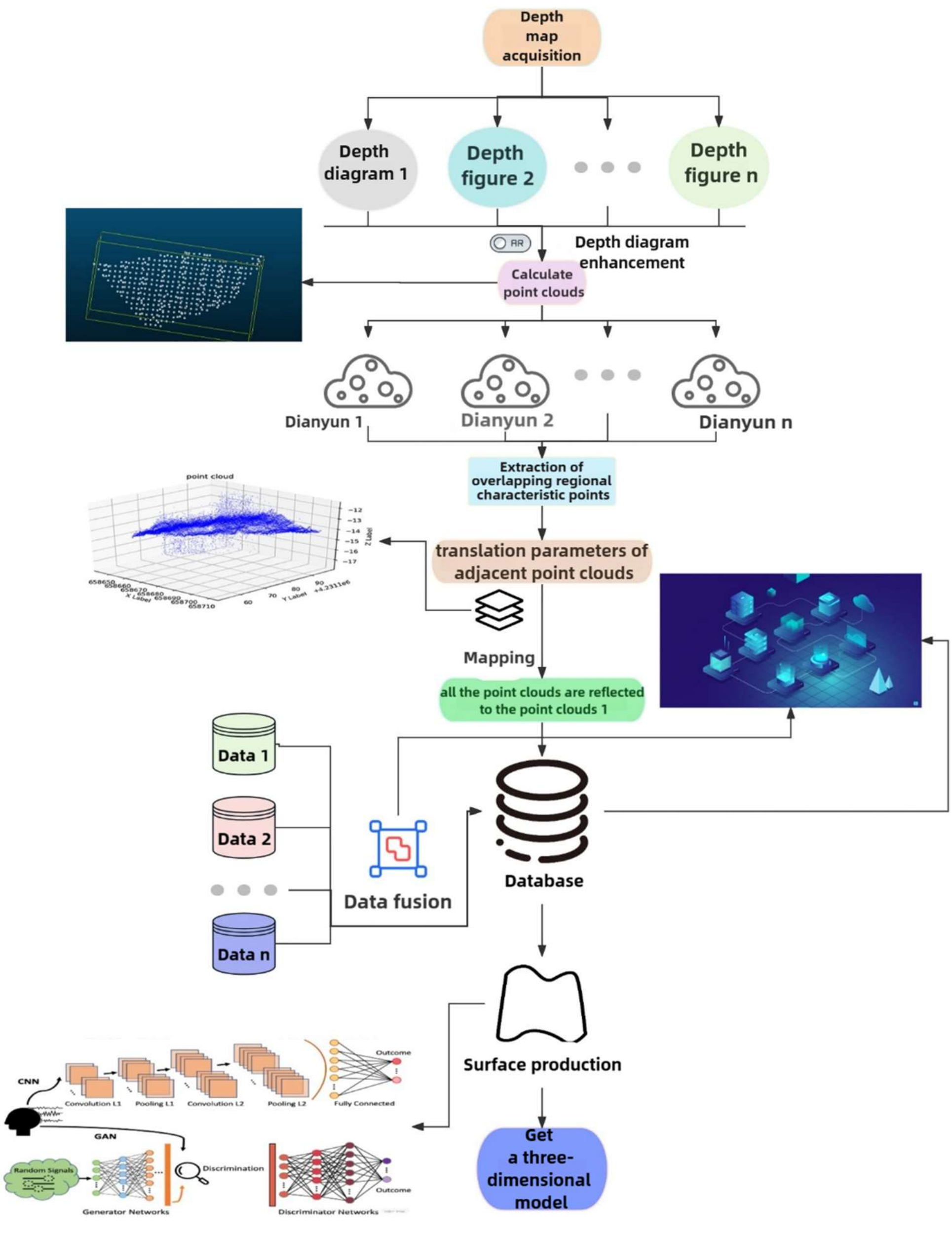

Fig. 5 3D reconstruction flow chart

ISSN: 2414-1895 DOI: 10.6919/ICJE.202503_11(3).0020

Usually full rank, SVD decomposition, the last column of the V matrix is the least square solution of Ax = 0, obtained by dividing the first 3 elements by the fourth element. Formulas (1) ~ (6) give the theoretical derivation of the pixel coordinates of the target point in different images and restore its specific position in the world coordinate system.

The system uses multiple high-resolution cameras to capture the environment and human body dynamics from different angles, ensuring that there are no blind spots in all directions, achieving multi-view images, and then data processing and analysis through integrated computer vision technology and deep learning algorithms. This 3D modeling technology allows the system to fully understand the specific position and volume of the human body in space, greatly improving the dimension and accuracy of monitoring. The system also performs human pose estimation, and by analyzing the human pose and movement in the reconstructed 3D model, the system can gain insight into more levels of behavioral information.

2.2 Research Status at Home and Abroad

Domestic research started later than the West, but after a period of development, face recognition, machine vision and related technologies have also developed rapidly. Smart agriculture combines modern information technology and intelligent agricultural machinery and equipment to form a new agricultural development model[4]. It utilizes advanced sensors, Internet of Things (IoT) devices, and big data analytics to monitor and manage various aspects of farming, including soil conditions, crop growth, and weather patterns. This allows farmers to make more informed decisions and optimize resource utilization, leading to increased efficiency and sustainability in agricultural production. Additionally, smart agriculture incorporates robotics and automation for tasks such as planting, harvesting, and sorting, reducing labor costs and improving productivity. Overall, the integration of technology into agriculture is transforming traditional farming practices and paving the way for a more modern and advanced agricultural sector. The detection of small target defect on strip surface is a research hotspot in the field of industrial quality inspection. Small and medium-sized target defects are easy to be missed in surface defect detection of hot rolled strip steel[5]. The improved yolo v7 algorithm is more accurate in the detection scene defect detection of small target hot rolled strip. Health monitoring in the field of civil engineering is of great significance to ensure the long-term and stable service of the project. Compared with traditional monitoring methods, computer vision technology based on deep learning has the advantages of high efficiency and accuracy[6]. At the same time, in the field of civil engineering, computer vision technology based on deep learning also plays a great role in construction site safety management, local damage detection of service structures and overall damage assessment of post-disaster structures[7].

Some foreign research institutions have made important contributions to the research of computer vision technology based on YOLO algorithm. YOLO, or You Only Look Once, is a real-time object detection system that is widely used in the field of computer vision due to its high efficiency and accuracy. These research institutions have promoted the development of computer vision technology in many fields by deeply exploring the optimization and innovation of YOLO algorithm. For example, in the field of medical research, in vitro micronucleus testing is a standard genotoxic test used to detect the potential genotoxic effects of chemicals or physical factors on cells. By applying the YOLO algorithm to data analysis for this type of experiment, researchers can identify and analyze micronucleus more quickly and accurately, improving the efficiency and reliability of the experiment. This interdisciplinary collaborative research not only demonstrates the powerful potential of YOLO algorithm, but also provides new ideas and methods for genotoxicity research. Face recognition is widely used in today's society, but it is difficult to accurately identify faces. In-depth learning through joint recognition and verification can detect and recognize faces accurately and quickly[8]. There are also deep face recognition in foreign countries, and the VGGface model of Cambridge University has laid a deep foundation for face recognition[9]. With the development of camera technology, the appearance and application of face recognition cameras have enriched the application environment of face recognition[10].

ISSN: 2414-1895 DOI: 10.6919/ICJE.202503_11(3).0020

3. Specific Goal

On the one hand, it is to achieve human-machine separation, so as to improve operational safety and efficiency. The achievement of this goal means that in many high-risk work environments, human employees can complete their jobs by remotely controlling robots or other automated devices, away from potential hazards. This not only greatly improves safety, but also reduces human error and increases productivity. On the other hand, through computer vision and deep learning, this study processes the image data captured by multiple cameras, so that the system not only recognizes the existence of the human body, but also identifies the key feature points of the human body, and insights more levels of behavioral information, laying the foundation for future research. This technological advancement enables monitoring systems to more accurately understand and analyze human behavior, which is of great significance for both security and behavioral research. The project of multi-intelligent security monitoring and space electronic fence in the work area not only has important scientific and technological value, but also carries far-reaching social impact and responsibility. Through the integration of advanced monitoring technology and intelligent analysis systems, the project can provide real-time and comprehensive safety monitoring in a variety of work environments, thereby effectively preventing accidents. At the same time, through the space electronic fence technology, access control can be carried out in specific areas to further ensure the safety of the workplace. Through the implementation of this project, we expect to make positive contributions to social security, smart city construction and scientific and technological progress. We hope that this technology can play an important role in the construction of smart cities, improve the overall safety management level of cities, and create a safer and more convenient living environment for residents. In addition, the research will also promote technological innovation and industrial development in related fields, injecting new vitality into economic growth.

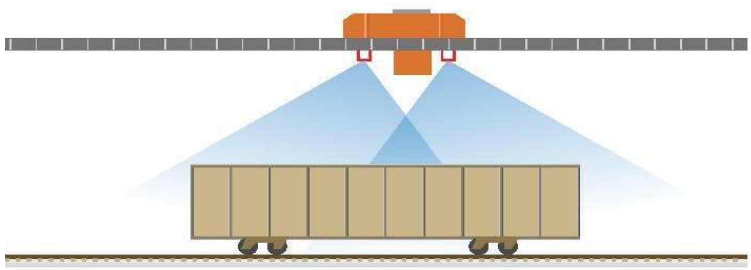

Fig. 6 Cameras installed on the front and rear of the car

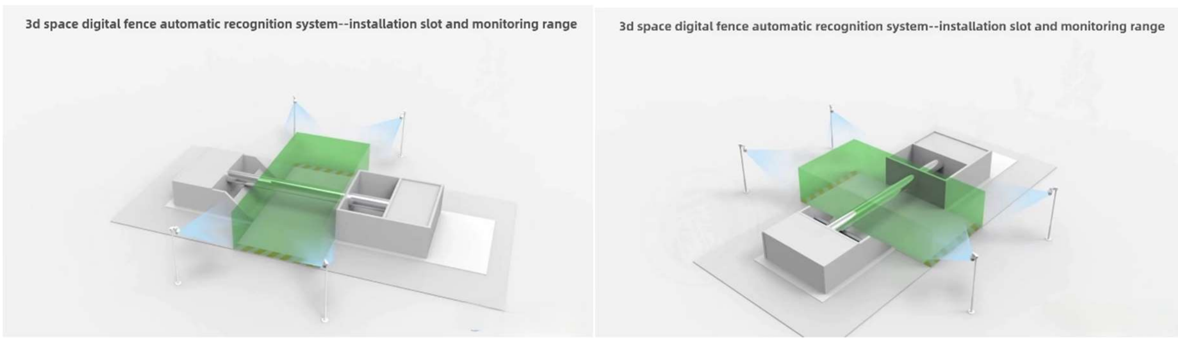

Fig. 7, 8 Electronic fence automatic identification system

ISSN: 2414-1895
DOI: 10.6919/ICJE.202503_11(3).0020

4. Conclusion

Through the research and development of intelligent isolation driving system, it can effectively isolate personnel and mechanical operations, significantly reduce the incidence of safety accidents in the process of operation, and ensure the life safety of staff. By monitoring and tracking the personnel dynamics in the working area in real time, the intelligent isolated driving system ensures the smooth progress of the working process, reduces the work interruption caused by safety accidents, and thus improves the overall working efficiency and production efficiency. This system not only breaks through the limitation of traditional monitoring means, but also significantly improves the ability of real-time perception and accurate analysis of personnel dynamics and environmental changes.

Acknowledgments

We would also like to thank the institutions for their facilities and technical resources. Special thanks to Editor Wang for his assistance on specific tasks, and to Teacher Ma for his valuable insights into the research. Finally, we would like to thank the anonymous reviewers for their constructive comments, which have greatly improved the quality of this paper.

References

[1] Xie Shuping, Yang Bo, WAN Xinjun, HUANG Pengcheng, JIANG Jianyu. Three-dimensional vibration measurement technology based on binocular stereo vision [J]. Optical Instruments,2018,40(5):15-21. (in Chinese).

[2] Zhang Yanwen; Hu Kai; Wang Pengsheng. Review of 3D reconstruction algorithms [J]. Journal of Nanjing University of Information Science and Technology (Natural Science Edition),2020(05).

[3] Zhang Hongyu; Jiao Xinquan; Jia Xingzhong. Design of binocular vision acquisition and processing system. Fire & Command & Control,2023(07).

[4] Qin Tao, Zhuang Weidong. Research status of machine vision technology in modern agricultural production [J]. Modern Agriculture, 2024, (04): 90-93.

[5] Fan Rong, Ma Xiaolu. Improved YOLOv7 algorithm for Small target Defect Detection on Strip Surface [J]. Journal of Hefei University of Technology (Natural Science Edition), 2024, 47 (03): 303-308.

[6] Fang Cheng, YU Shengxin, LI Yonggang, JIA Wanglong, Yang Pengbo, Yang Xinyue. Computer vision health monitoring in Civil engineering based on Deep learning [J]. Journal of Tongji University (Natural Science), 2024, 52 (02): 213-222.

[7] Li Tao, LIU Junjiang, Zhang Cong, ZHU Lei. Metal surface defect detection based on Deep learning [J]. Journal of Normal Science, 2024, 44 (01): 36-42.

[8] Sun Yi, Wang Xiaogang, Tang Xiaoou. "Deep learning face representation through joint recognition verification." Advances in Neural Information Processing Systems, 1988-1996.

[9] Parkhi, Omkar M., and Andrea Vedaldi. "Deep face recognition." University of Oxford (2015).

[10] Chen Ruwei. Design of Automatic Adjustment Transmission Mechanism for Face Camera [J]. China Machinery, 2023, (26): 10-13.