Elevator Passenger Abnormal Behavior Recognition Method Based on Digital Twin

Xiaolin Li and Yaohui Song*
School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Keywords: Digital Twins, Posture Recognition, Bone Extraction, YOLOv7-OpenPose.

Abstract: Aiming at the lack of abnormal behavior dataset and scarcity of samples of elevator passengers, a method based on digital twin is proposed by this paper to build a vertical elevator passenger abnormal behavior detection platform and realize the virtual and real mapping of elevator operation status and passenger behavior. The digital twin scene is combined with the theory of human behavior modeling to enhance the abnormal behavior of passengers and provide sufficient abnormal behaviour data sources. In order to solve the problem of confusion between passengers and car background caused by the small range of elevator monitoring and reduce the accuracy of feature extraction, YOLOv7-OpenPose is used by this paper to obtain human bone features, which improves the recognition accuracy on the premise of ensuring the recognition speed, and realizes the rapid recognition of passengers' abnormal behaviors fused with twin data. Experimental results show that the proposed method not only demonstrates the feasibility, efficiency and security of digital twin technology in the creation of abnormal data, but also reflects the superiority of the improved algorithm in pose recognition.

1 INTRODUCTION

Elevator safety is closely related to public safety, according to the public data of the State Administration for Market Regulation on national special equipment, by the end of 2022, the number of elevators in China has reached 9.6446 million. Due to the small and closed space of the elevator car, the passenger dynamics cannot be controlled, and a series of safety problems such as door opening, falling, and blocking of cameras are prone to occur, so the research on the abnormal behavior of elevator passengers is of great significance.

At present, there is an endless stream of research on the abnormal behavior of elevator passengers. Lv et al. (Lv, 2021) used the YOLOv3 algorithm and the AlphaPose algorithm to collect and extract elevator passenger behaviors, and used SVM and neural networks to classify different abnormal behaviors. Shi et al. (Shi, 2021) used the OpenPose algorithm to extract the key nodes of the key frame sequence, obtain the spatial information of the target behavior, and identify the human behavior in the elevator car. Feng et al. (Feng, 2021) proposed a detection method of machine vision and multi-feature fusion for elevator passenger falls. Wang et al. (Wang, 2018) used Lucas-Kanade optical flow to design a passenger abnormal behavior detection system in the elevator car based on video recognition technology. Reinsalu et al. (Robal, 2023) proposed a method for identifying unsafe behaviors of car passengers based on deep learning and a fault warning method for brakes and control cabinets based on infrared spectrum analysis to identify abnormal behaviors online. Yu and Sun et al. (Yu, 2020), (Sun, 2019) proposed abnormal behavior detection models based on optical flow method and angular kinetic energy for the fighting behavior in the elevator car, respectively.

However, in the process of model training, testing, and validation, the datasets in the above studies are insufficient. The traditional elevator abnormal behavior sample data came from real human movements and the construction of human physical models (Yuan, 2020), (Fuller, 2023), respectively. For datasets created with real human actions, the execution process has security risks and requires high time and equipment costs. For the construction of human body physical models, the model has fixed actions and lacks real-time and

Song, Y. and Li, X.
Elevator Passenger Abnormal Behavior Recognition Method Based on Digital Twin.
DOI: 10.5220/0012887900004536

In *Proceedings of the 1st International Conference on Data Mining, E-Learning, and Information Systems (DMEIS 2024)*, pages 87-93
ISBN: 978-989-758-715-3

flexibility. The emergence of digital twins provides a solution to the above problems. The interaction between the digital twin model and the digital twin data (Saravanan, 2022) can not only realize the driving and updating of the digital twin model, but also support the storage, update and fusion of the digital twin data. On the one hand, it can realize the real human-computer interaction process (Amara, 2023), so as to realize the real-time monitoring of the scene in the elevator. And on the other hand, it can use the high-fidelity twin scene to reproduce the abnormal behavior of various passengers, which can not only ensure the safety of personnel and property and save equipment costs, but also provide sufficient data sets for model testing and verification for the expansion of sample data.

In summary, a method based on digital twins is proposed by this paper to establish a vertical elevator passenger behavior detection model. This method builds a digital twin of elevator passenger behavior monitoring and a human behavior twin model, and enhance the abnormal behavior data of elevator passengers. It provides a large amount of twin data for the identification of abnormal behaviors of elevator passengers. For the recognition of abnormal behaviors, this paper uses the YOLOv7-OpenPose algorithm to extract features and iteratively learn from the twin data, so as to realize the efficient recognition of passengers' abnormal behaviors.

2 CONSTRUTION OF VERTICAL ELEVATOR PASSENGER BEHAVIOR MONITTORING PLATFORM

2.1 Design of a Digital Twin Framework for Elevator Passenger Behavior Monitoring

Referring to the five-dimensional model architecture of digital twin proposed by He et al. (He, 2020), this paper builds a framework for the digital twin vertical elevator passenger behavior monitoring platform, as shown in Figure 1, which is divided into four layers, namely the physical layer, the interaction layer, the virtual layer and the application layer.

(1) Physical layer: The physical layer contains all kinds of hardware physical entities for elevator operation, and the main hardware equipment includes elevator car, camera in the car, elevator attitude sensor, etc., which are combined into scene monitoring entities, and ordinary sensors are responsible for collecting elevators static data, such as geometric dimensions, car status, etc., attitude sensors are responsible for collecting passenger behavior information.

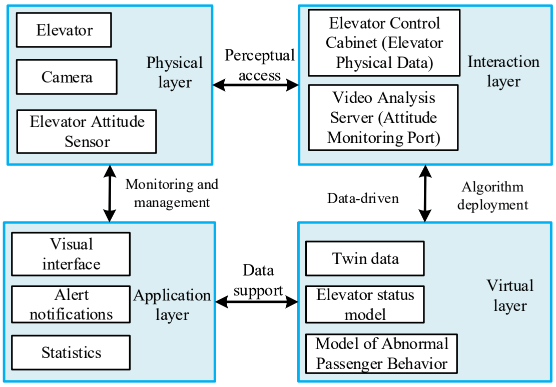

Figure 1: Elevator passenger behavior monitoring digital twin framework.

(2) Interaction layer: The interaction layer is mainly composed of elevator control cabinet and video analysis server. As the connection layer between the physical entity and the twin, the interaction layer can not only use the edge controller to convey the instructions of the elevator system, but also upload the simulation data to the video analysis server for model training, and at the same time transmit all kinds of collected data to the virtual layer to drive the call of the twin model and the interaction of the twin data.

(3) Virtual layer: As the data combination of physical entities, the virtual layer contains all the space, attributes and management data of the physical layer, and stores, transmits, expresses and deepens them, mainly by means of new technologies such as digital modeling, Internet of Things, and artificial intelligence. The twin model is mainly composed of an elevator state model and a human behavior model. With the help of virtual and real interaction technology, the elevator state model integrates real-time information of elevator equipment on the basis of the digital model to realize the real-time monitoring of the operation status of elevator equipment. The human behavior model uses video data to realize the virtual and real mapping of real passenger human behavior, and at the same time, the digital twin character model creates a large amount of abnormal behavior data. The twin data comes from the physical layer and the virtual layer, the data of the physical layer is used to drive the twin model, and the data supply of the virtual layer is called by the layer and the interaction layer.

(4) Application layer: The application layer directly serves the staff, including the visualization of the twin interface and the operation functions of safety warning, and displays the elevator status and the elevator passenger status in real time, so as to realize the three-dimensional dynamic supervision of elevator equipment and passenger behavior. It mainly uses UI design and three-dimensional visualization technology, and connects to the communication system at the same time, so that elevator passengers can take rescue actions in the first time in case of safety accidents.

2.2 Construction of Digital Twin Scene for Passenger Behavior Monitoring

The vertical elevator passenger behavior monitoring digital twin platform was built in Unreal Engine 5 (UE5 Unreal Engine), as shown in Figure 2, which is divided into three steps: geometric model drawing, twin scene construction, and virtual and real mapping.

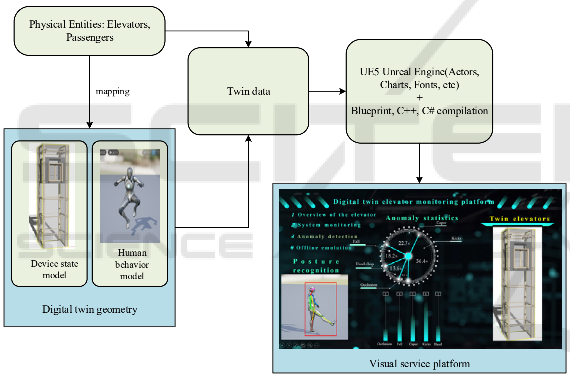

Figure 2: Construction of digital twin scene for elevator passenger behavior monitoring.

2.2.1 Geometric Model Drawing

The geometric model is drawn according to the physical entity behavior rules and each system, and the elevator model is drawn by SolidWorks, including eight systems: drag system, car, weight balance system, power system, door system and safety protection. The character model was drawn in Maya and had to be drawn with a skeletal mesh to reflect the real human behavior.

2.2.2 Twin Scene Building

Import the above geometry into UE5 and assemble it according to the inclusion relationship in the physical world. When building a scene, pay attention to redefining material properties, hierarchical relationships, motion constraints, and so on. Ensure that the movement of the twin in the twin scene is consistent with the physical world.

2.2.3 Virtual-Real Data Mapping

Twin data is the key to real-time mapping of physical entities and twins. In this paper, the elevator status data is obtained through the RS485 communication bus between the external interface board and the elevator control cabinet, and stored in the corresponding MySQL data for driving the twin. Passenger behavior data includes video data and passenger attitude information, wherein passenger attitude information is obtained by using the YOLOv7-OpenPose algorithm proposed in this paper on the video analysis server, and sent to the client through TCP network protocol and Socket communication, and the client can realize the construction of visualization platform services by using C++ and BluePrint to parse and distribute the input data.

3 IDENTIFICATION OF ABNORMAL PASSENGER BEHAVIOR FUSED WITH TWIN DATA

3.1 Passenger Abnormal Behavior Modelling and Data Augmentation

Abstracting the human skeleton into a collection of several bones and joint points can not only reduce the complexity of constructing a complete human behavior model, but also the human bone structure is similar, and the human behavior can also be described using the position information of these joints. As shown in Figure 3, The traditional 3D coordinate description method makes the position parameters of each joint point independent of each other during the movement. It violates the constraint that the skeleton length of the mannequin remains unchanged, so this paper directly uses a hierarchical method to equivalence the mannequin to the joint tree shown in the figure above. The root node of the tree is used as the geometric center of the human skeleton model to control the overall displacement and direction of the model. The remaining sub-nodes are indirectly or directly connected to the root node to assist in the determination of the posture of the mannequin and the presentation of the movement process (Han, 2020).

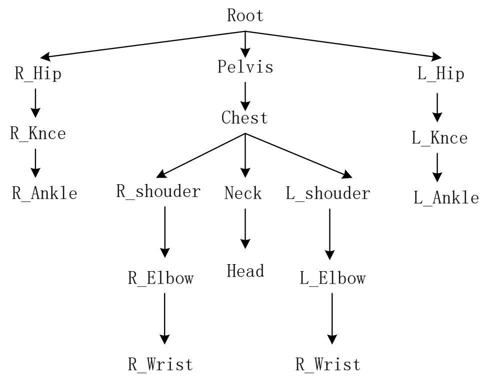

Figure 3: Human joint tree.

As shown in Figure 4, after equating the human skeleton to a joint tree, the movement of the whole human body can be described in detail by means of forward kinematics and inverse kinematics (Moharkan, 2023).

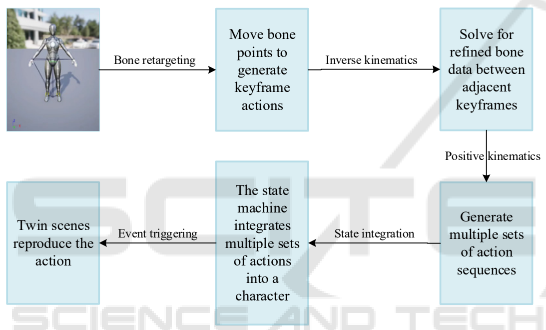

Figure 4: Digital twin-based passenger anomaly modelling process.

Among them, after the displacement and rotation of each joint relative to its parent joint are known, the posture of the whole human body after movement can be solved by using positive kinematics. With the initial and ending human posture known, inverse kinematics can be used to solve for specific changes in each joint during the movement process.

Through forward and reverse kinematic behavior modeling, various dangerous posture actions can be created for the twin model to operate, because the abnormal behavior recognition research of machine learning requires a large amount of video image data. In the process of acquisition, the frame rate output of the CameraActor is set to 30frame/s, the resolution is set to 1920×1080, the image output format is video sequence (avi), the angle is set to the horizontal plane of the parallel car roof, and the four virtual cameras are recorded at the same time through the control command. Data enhancement is carried out for abnormal behaviors such as jumping, falling, kicking, and hand-picking doors in the elevator.

3.2 Human Bone Detection Based on YOLOv7-OpenPose

Aiming at the lack of abnormal behavior cases and insufficient datasets in real life, the abnormal behavior modeling of passengers based on digital twins can provide a large number of training and test samples for abnormal posture and behavior. In this paper, we propose a passenger abnormal behavior recognition model fused with twin data, and the overall framework diagram of the model is shown in Figure 5.

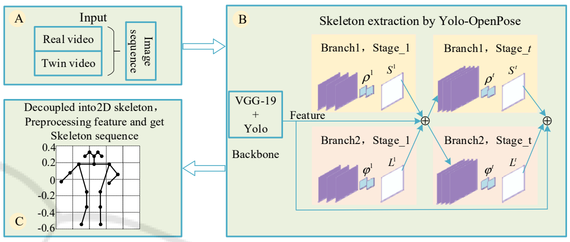

Figure 5: Bone extraction of passenger abnormal behabior fused with twin data.

The video data of passenger abnormal behavior collected in practice and the twin data of passenger abnormal behavior constructed in 2.1 were processed into image sequences, and the skeletal posture of the human body was extracted by the YOLOv7-OpenPose algorithm.

OpenPose, as a classical pose extraction algorithm, has a multi-stage, supervised convolutional neural network (Moharkan, 2023), which is widely used in human behavior recognition, and its overall structure is shown in B of Figure 5. First, the feature map is extracted through the backbone F . Secondly, it inputs this map into Branch1 and Branch2 of the first stage Stage_1. And then it obtains the 2D confidence map of the joint point S^1 and the partial affinity domain L^1 respectively. Finally, it makes F , S^{t-1} and L^{t-1} as the input of the next stage Stage_t ($t \ge 2$), and perform multiple iterations. The inputs and outputs of Branch1 and Branch2 at different stages are shown in Eq. (1).

$$\begin{cases} S^t = \rho^t(F, S^{t-1}, L^{t-1}), \forall t \ge 2 \\ L^t = \varphi^t(F, S^{t-1}, L^{t-1}), \forall t \ge 2 \end{cases} \quad (1)$$

, where ρ^t and φ^t , represent the network relationship between the 2D confidence graph and partial affinity domain of the t stage and the feature map, 2D confidence graph and partial affinity domain of the stage $t-1$, respectively. Through the continuous iteration of the above multi-stage convolutional neural network, the more joint points of the human body are obtained. For the position d_{j1} and d_{j2} of any two joint points, the confidence degree of the joint point pair needs to be characterized by calculating the linear integral of the partial affinity domain by Eq. (2). At last, it selects the joint with the highest confidence to complete the splicing, and gets the skeletal posture of the whole human body.

$$\begin{cases} E = \int_{u=0}^{u=1} L_c(p(u)) \times \frac{d_{j2}-d_{j1}}{\left\| d_{j2}-d_{j1} \right\|} d_u \\ p(u) = (1-u)d_{j1} + ud_{j2} \end{cases} \tag{2}$$

, where L_c represents a partial affinity domain for a pair of joint points.

Although the OpenPose algorithm has the advantages of fast recognition speed and high accuracy in large scenarios. However, the scope of elevator monitoring is small, which is easy to cause confusion between passengers and the car background and reduce the accuracy of feature extraction. So it is necessary to consider more dependencies between deep and shallow networks to capture more details and textures. VGG-19 (Wen, 2019) is the feature extraction network of the OpenPose algorithm, and the continuous iteration of the convolutional network makes the significance of the feature map continuously decrease. Therefore, in order to better extract the skeletal features of passengers in the car environment, this paper improves the traditional VGG-19 and proposes a skeletal point detection algorithm for elevator passengers based on YOLOv7-OpenPose. The backbone network structure of the algorithm is shown in Figure 6.

This structure uses the YOLOv7 object detection model to extract abnormal passenger targets. The YOLOv7 series uses both object detection and object classification, and its object detection layer of the 79th layer 13*13 detection network and the object detection layer of the 91-layer 26*26 detection network are both convolutional layers, and only feature extraction is performed on the target. The specific position, bounding box and confidence level of the target are obtained through the object detection network of YOLOv7, which is placed before the third Maxpool layer in the VGG-19 network as input. And then the output tensor of the whole object detection is spliced with the output tensor of the original Maxpool layer as the input tensor of the next layer and the bone extraction is continued.

4 EXPERIMENTAL RESULTS AND ANALYSIS

4.1 Experimental Environment

This experiment adopts Unreal Engine 5.3.2 under Windows 10 operating system, TensorFolw framework, Core i9-10980XE processor, 128G

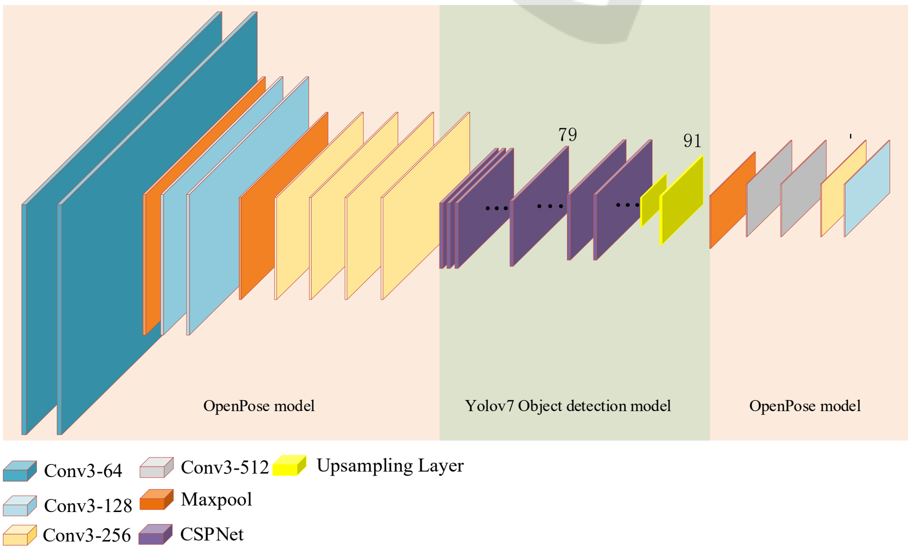

Figure 6: YOLOv7-openpose algorithm backbone.

memory, Core i7-11400F processor, 32G memory, Nvidia RTX 3080 graphics processor and Nvidia RTX3060 graphics processor.

4.2 Dataset

The experiments were performed on a custom elevator passenger abnormal behavior dataset and a COCO2017 dataset. Among them, the COCO2017 dataset is used to test the YOLOv7-OpenPose bone point detection algorithm in this paper. The dataset contains a total of 163957 images of the training, validation, and test sets, and each image is labeled with 18 bone points of the human body.

By capturing and merging the videos in the real scene and the twin scene at a rate of 30 frames per second, the custom passenger abnormal behavior dataset shown in Table 1 is formed.

Table 1: Customize the Passenger Abnormal Behavior Dataset.

Action type	The amount of real data	The amount of virtual data	Total
Caper	28	2023	20510
Hand chop	47	2028	2075
Kicks	66	2047	2113
Fall	20	2011	2031
Total	161	8109	8270

4.3 Human Skeletal Point Detection Analysis

Human skeleton detection based on YOLOv7-OpenPose is not only the key link to obtain passenger posture information in the real environment, but also the core content of obtaining passenger abnormal behavior dataset. For the abnormal behaviors of various passengers constructed in the twin scene, the skeletal point detection results are shown in Figure 7.

4.4 Comparison of the Performance of the YOLOv7-OpenPose Algorithm

In order to verify the accuracy and real-time performance of the YOLOv7-OpenPose bone detection algorithm, the performance of the text was compared with the classical OpenPose algorithm on the COCO dataset, and the performance of different series of YOLO algorithms was analyzed. The experimental results are shown in Table 2.

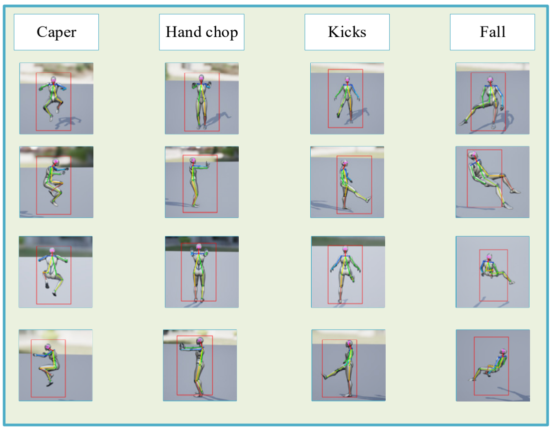

Figure 7: Bone diagnosis of abnormal passenger behavior based on twin data.

Table 2: Comparison of the performance of different networks on the COCO test set.

The type of algorithm	OpenPose	OpenPose + YOLOv5	OpenPose + YOLOv7	OpenPose + YOLOv8
AP	0.701	0.728	0.734	0.741
AP^{50}	0.854	0.871	0.882	0.845
AP^{75}	0.711	0.722	0.736	0.709
AP^{M}	0.695	0.714	0.719	0.725
AP^{L}	0.774	0.776	0.779	0.785
Time-consumed/s	21341	23448	22154	25243

Where AP is the mean index of average accuracy. AP^{50} and AP^{75} represent the joint point predictors with thresholds of 0.50 and 0.75, respectively, AP^{M} and AP^{L} represent the predictors of small and large human body size, respectively. Time-consumed is the test time consumed by the algorithm on the experimental platform in this paper, which is used to characterize the algorithm complexity. Comparing the accuracy and time complexity of each model in Table II, it can be seen that the AP^{50} and AP^{M} prediction indicators of YOLOv7-OpenPose perform the best among the above algorithms. Compared with the original model OpenPose, the accuracy indicators of YOLOv7-OpenPose are improved and the amplitude is about 4%, and the introduction of the YOLOv7 module does not have a great impact on the time complexity of the overall YOLOv7-OpenPose algorithm. The average processing time of the algorithm is still close to that of the original OpenPose algorithm, which can meet the real-time requirements.

5 CONCLUSIONS

In order to solve the problem of lack of a large number of samples in the research on abnormal behavior recognition based on machine learning caused by the scarcity of abnormal behavior data of elevator passengers, this paper builds a digital twin scenario for monitoring abnormal behavior of passengers in vertical elevators. Based on the human skeleton model and kinematics principle, the abnormal behavior data was constructed, and a total of 8270 twin actions were provided for abnormal behavior recognition, and finally the improved YOLOv7-OpenPose human skeleton detection algorithm was used. The experimental results show that the accuracy of the model is improved by about 4% on the basis of the original model OpenPose, and the model does not significantly increase the time complexity in terms of real-time, which solves the problem of low feature extraction rate caused by the confusion between passengers and car background and the trade-off between real-time and accuracy. The modeling and recognition of abnormal behaviors proposed in this paper have the characteristics of high accuracy, strong real-time performance and good interactivity. In the future, further research will be carried out on the abnormal behavior of multiple people in complex scenes in the elevator.

REFERENCES

Lv, X., Wang, S. and Chen, T., 2021. Human gait analysis method based on sample entropy fusion AlphaPose algorithm. *2021 33rd Chinese Control and Decision Conference (CCDC).* pp.1543-1547. IEEE. doi: 10.1109/CCDC52312.2021.9602427.

Shi, D., Zhu, W. and Cheng, R., 2022. Human Fall Detection Algorithm Based on YoloX-s and Lightweight OpenPose. *2022 15th International Conf. on Advanced Computer Theory and Engineering (ICACTE).* pp. 23-28. IEEE. doi: 10.1109/CCDC52312.2021.9602427.

Feng, S., Niu, K. and Liang, Y., 2021. Research on elevator intelligent monitoring and grading warning system. *IEEE International Conf. on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI).* pp. 145-148. IEEE. doi: 10.1109/CEI52496.2021.9574579.

Wang, Z., Yang, X., 2018. Moving target detection and tracking based on Pyramid Lucas-Kanade optical flow.*2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC).* pp. 66-69. IEEE. doi: 10.1109/ICIVC.2018.8492786.

Reinsalu, U., Robal, T., 2023. A Touch-Free Service Button for Smart Elevator Operation with Dynamic QR-code Generation. *2023 12th Mediterranean Conference on Embedded Computing (MECO).* pp. 1-4. IEEE. doi: 10.1109/MECO58584.2023.10155102.

Yu, L., Wang, Y. and Li, J., 2020. Algorithm of Detecting Abnormal Behavior Based on Optical Flow Method. *2020 2nd International Conference on Information Technology and Computer Application (ITCA).* pp. 191-193. IEEE. doi: 10.1109/ITCA52113.2020.00047.

Sun, Z., Xu, B. and Lu, M., 2019. A real-time video surveillance and state detection approach for elevator cabs. *2019 International Conference on Control, Automation and Information Sciences (ICCAIS).* pp. 1-6. IEEE. doi: 10.1109/ICCAIS46528.2019.9074707.

Yuan, R., Zhang, Z. and Song, P., 2020. Construction of virtual video scene and its visualization during sports training. *IEEE Access*. vol. 8, pp. 124999-125012. IEEE. doi: 10.1109/ACCESS.2020.3007897.

Kyrollos, D., Fuller, A. and Greenwood, K., 2023. Under the cover infant pose estimation using multimodal data. *IEEE Transactions on Instrumentation and Measurement*. vol. 72, no. 5007212, pp. 1-12. IEEE. doi: 10.1109/TIM.2023.3244220.

Saravanan, S., Muthusenthil, B. and Gurusubramani, S., 2022. A review of digital twin leveraging technology, concepts, tools and industrial applications. *2022 1st International Conf. on Computational Science and Technology (ICCST)*. pp. 742-748. IEEE. doi: 10.1109/ICCST55948.2022.10040359.

Amara, K., Kerdjidj, O. and Ramzan, N., 2023. Emotion Recognition for Affective human digital twin by means of virtual reality enabling technologies. *IEEE Access*. vol. 11,pp.74216-74227,doi:10.1109/ACCESS.2023.3285398

He, Z., Yuan, T. and Li, X., 2022. Research on Data Transfer Scheme of Virtual-Real Interactive in Five-Dimensional Digital Twin System. *2022 IEEE 8th International Conf. on Computer and Communications (ICCC).* pp. 2339-2344. IEEE. doi:10.1109/ICCC56324.2022.10066033.

Han, B., Jiao, Y. and Liu, G., 2020. Kinematic & dynamic models of human lower extremity during the gait cycle. *2020 6th International Conference on Control, Automation and Robotics (ICCAR).* pp. 568-573. IEEE. doi: 10.1109/ICCAR49639.2020.9108090.

Moharkan, M. J., 2023. Comprehensive Survey on Human Motion Tracking System. *2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS).* pp. 1380-1387. IEEE. doi: 10.1109/ICACRS58579.2023.10405344.

Wen, L., Li, X., 2019. A new transfer learning based on VGG-19 network for fault diagnosis. *2019 IEEE 23rd International Conf. on Computer Supported cooperative work in design (CSCWD).* pp. 205-209. IEEE. doi: 10.1109/CSCWD.2019.8791884.