IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 1, 2020, accepted July 19, 2020, date of publication July 27, 2020, date of current version August 5, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3011951

Optimization of Wireless Video Surveillance System for Smart Campus Based on Internet of Things

ZHIQING ZHOU, HENG YU, AND HESHENG SHI
School of Information Engineering, Pingdingshan University, Pingdingshan 467000, China

Corresponding author: Zhiqing Zhou (2544@pdsu.edu.cn)

This work was supported in part by the Science and Technology Research Plan of Henan Province (Research on Intelligent Physical Signs Monitoring System of Home-Based Elderly Care Based on IoT Technology) under Project 162102310248, and in part by the Pingdingshan University Fund Carry Forward Project (Research on Key Technologies of Intelligent Home Care Equipment for the Elderly Based on IoT Technology) under Project JZ2017026.

ABSTRACT In order to strengthen school security and build a wireless smart campus, this article optimizes the existing wireless video surveillance system based on the Internet of Things. This paper first optimizes the surveillance quality in the video surveillance system, and proposes a zero-copy buffer strategy, a network congestion suppression strategy, and a codec rate coordination strategy. Secondly, for the distributed wide area video surveillance system, a tracking optimization method based on multi-camera fusion is proposed. Finally, this paper constructs a Bayesian monitoring event modeling method based on genetic algorithm. Experimental results show that the optimized video surveillance system has basically stable delay, significantly reduced packet loss rate, and smooth video playback. This method can effectively realize the coordinated tracking of multiple cameras in a wide-area monitoring scenario, achieve high tracking and monitoring performance, and meet the requirements of smart campus construction.

INDEX TERMS Internet of Things, smart campus, video surveillance, optimization.

I. INTRODUCTION

With the acceleration of the construction of colleges and universities, the campus area is gradually expanding, and the number of mobile personnel is also increasing. Various uncertain factors have also brought great difficulties to the safety management of the campus. Video surveillance is an important means to maintain a safe campus. However, there are still many analog and digital hybrid systems used in the video surveillance system of colleges and universities, which can no longer meet the requirements of the development and construction of modern digital campuses and smart campuses [1]. Therefore, building a digital, networked, and intelligent video surveillance system is not only to meet the requirements of campus security, but also an important part of a smart campus. Internet of Things (IoT) [2]–[5] as an important part of the new generation of information technology, through radio frequency identification (RFID), infrared sensors, global positioning system (GPS), laser scanners and other information sensing equipment, according to the agreed agreement connect any items in the sensor network to the Internet, exchange information and communicate, and realize the intelligent identification, positioning, tracking, monitoring and management of items. The characteristics and functions of IoT make it widely used in smart homes, smart transportation, industrial monitoring, and other fields [6]–[8].

As an important security means, the video surveillance system has a long history of development, and it has been widely used in the fields of public safety monitoring, production process monitoring, and residential community monitoring [9]. With the further development of multimedia technology and the gradual improvement of people's requirements for quality of life, research on video surveillance systems has received great attention. On the one hand, the development of image processing technology and the level of communication technology have driven the development of the technology of video surveillance systems. In turn, the technical level of video surveillance systems has also reflected from one side the development of image processing technology and the

The associate editor coordinating the review of this manuscript and approving it for publication was Zhihan Lv.

development of communication technology and its industrialization level. Foreign countries have begun to adjust the video bit rate through adaptive rate control technology, and the obtained video quality optimization has also been widely recognized. Eswara *et al.* [10] proposed a video rate adaptation method. This method can achieve low latency of adaptive video streaming. Yahia *et al.* [11] evaluated the cost of the algorithm and combined the low-latency HTTP streaming technology with the low-latency video encoding technology to improve user perception. Muzaffar *et al.* [12] proposed a dynamic rate switching method for live video streams. Although it can detect the drop in throughput in time and adjust the sending bit rate accordingly to reduce the interruption of video playback, he ignores the frame skip caused by the throughput overflow. Pu [13] proposed a video playback threshold calculation method for throughput overflow by analyzing video. This method requires that the code rate is fixed and the threshold is calculated in advance. If the transmission bandwidth and the code rate change, the calculated threshold will no longer be optimal. Kesavan and Jayakumar [14] added an overflow threshold and an underflow threshold to reduce the occurrence of video playback interruptions and frame skips caused by changes in network bandwidth. The algorithm uses a proportional differential controller, which can accurately adapt to changes in the video bit rate, but the buffer size needs to be changed frequently.

Video target tracking technology is still facing many challenges in practical applications, such as target posture changes, scene lighting effects, rapid target movement, target occlusion disappearing, complex backgrounds, etc. [15]. Compressive tracking (CT) [16] algorithm is a simple and efficient target tracking algorithm based on compressed sensing, and is also a typical representative of pattern classification algorithm. Similar to the framework of the general pattern classification algorithm, the main idea of the CT algorithm is to first extract image features, and then use the classifier to classify the image, and finally use the decision result to update the classifier online. Tracking learning detection (TLD) [17] algorithm is an efficient long-term target tracking algorithm. Although this method can improve the tracking effect in some cases, it requires an offline learning process. Before detection, a large number of target samples are required for learning and training. This means that the training samples should cover the various shape changes, posture changes, and lighting changes that may occur to the target. Struck algorithm [18] questioned the positive and negative sample selection mode of the traditional algorithm, and innovatively proposed an adaptive visual target tracking algorithm framework based on structural output prediction results. Support vector machine is used as the classifier, and the tracking function is realized by explicitly introducing the output space, so as to avoid the intermediate classification link and directly output the tracking result. It should be noted that the algorithm does not output the predicted label, but outputs various transformations of the target, and then selects the optimal transformation as the final result. The particle filter algorithm [19] is a process of obtaining a minimum variance distribution by finding a set of random samples propagating in the state space to approximate the probability density function and replacing the integral operation with the sample mean.

Video target tracking based on multi-camera monitoring system [20] on the basis of solving the problem of single-camera target tracking, there are still many new problems that need to be faced, such as the deployment of cameras, the confirmation of target matching between cameras, and the transfer of camera observation data and fusion. Among them, target matching between cameras, that is, how to determine whether the targets observed in different cameras come from the same target, is the basis for continuous target tracking, and is also the difficulty of multi-camera target tracking. The same target, in the field of view of different cameras, the appearance, and posture of the target may be different. These problems pose challenges to target confirmation. Zhang and Cohen [21] borrowed from the related applications of wireless sensor networks and proposed a master-slave camera tracking scheduling algorithm. Osamy *et al.* [22] proposed and adopted a scheduling algorithm based on camera clusters, which overcomes the problem of tracking failure due to communication failure between cameras. Yang *et al.* [23] innovatively proposed a camera scheduling framework based on game theory. Quintana *et al.* [24] proposed a multi-camera collaborative algorithm based on Markov decision process. This method proposes a multi-camera scheduling framework based on decision theory. The framework aims to maximize the number of observation targets while ensuring the resolution of each monitoring target. Wang *et al.* [25] proposed a multi-camera scheduling method based on ant colony algorithm. This method uses the target position as the basis for camera scheduling. The camera priority, the distance between the target and the camera, and the visibility of the target are used to calculate the evaluation function of the target camera pair through the ant colony algorithm, and the optimal target camera pair is selected. He *et al.* [26] proposed a distributed camera network control method based on game theory. This method establishes a camera control theory based on game theory to ensure the coverage of the visual field.

In the aspect of intelligent video surveillance, my country started late, but its development is relatively rapid. Zhou [27] adopted the method of segmentation and marking of salient features of targets in video surveillance to realize the recognition and tracking of moving targets in video surveillance. He *et al.* [28] used the method of target segmentation and matching module to perform real-time tracking of targets in video surveillance. Choi and Na [29] studied a tracking system based on closed-circuit television. The system is based on a personal computer design, using a data acquisition card and equipped with automatic tracking software. It is necessary for the monitoring personnel to manually select the daily target to be tracked to achieve target tracking in the video monitoring screen. Taylor *et al.* [30] developed a personal computer-based highway gimbal monitoring system that applies video target detection and tracking technology to the highway monitoring system to perform traffic condition detection and monitoring. Iguernaissi *et al.* [31] developed a multi-camera joint intelligent monitoring system. The system consists of a panoramic camera and multiple area cameras. The panoramic camera is responsible for large-scale target detection. When a target is found and the area where the target is located is sent, an instruction is sent to the area camera. The regional camera is responsible for tracking the daily standard. When the target is about to leave the camera's field of view, it will feed back the instruction to the panoramic camera, and then the panoramic camera will be uniformly scheduled.

In order to better strengthen school security precautions, this article optimizes the existing wireless video surveillance system based on IoT. Specifically, the technical contributions of this article can be summarized as follows:

First: Firstly, the surveillance quality in the video surveillance system is optimized, and a zero-copy buffer strategy, network congestion suppression strategy and codec rate coordination strategy are proposed. The zero-copy buffer strategy reduces the terminal load and improves the system's processing power. The network congestion suppression strategy effectively reduces the packet loss rate. The codec rate coordination strategy balances system delay and fluency.

Second: Aiming at distributed wide area video surveillance system, a tracking optimization method based on multi-camera fusion is proposed. The tracking optimization algorithm performs data weighted fusion based on the priority of the target and the occlusion state of the target in each camera and the size of the segmented image, assigns high priority targets to the camera with the best weight for tracking, and dynamically balances the tracking load, assign the low priority target in the camera with heavy tracking load to other cameras for tracking.

Third: This paper builds a Bayesian monitoring event modeling method based on genetic algorithm. This method can achieve the purposes of reducing system redundancy, reducing optimization complexity, improving system modeling flexibility, reducing the amount of data required for optimization, and ensuring detection quality.

The rest of this article is organized as follows. Section 2 analyzes related concepts. Section 3 constructs IoT-based smart campus wireless video surveillance system optimization method. Section 4 carried out simulation experiments. Section 5 summarizes the full text.

II. RELATED CONCEPTS

A. INTERNET OF THINGS

IoT is based on the computer Internet, using RFID, wireless data communication and other technologies. IoT has three main characteristics:

(1) Comprehensive perception: perception is the core of IoT, and things are connected.

(2) Reliable transmission: real-time and accurate data transmission through the agreed unified communication protocol.

(3) Intelligent processing: the purpose of IoT is to realize intelligent identification, positioning, monitoring, and management of various items through emerging technologies such as cloud computing and artificial intelligence.

The IoT enables the application of public networks to be further extended and expanded from information exchange between people to information exchange and information processing between things. Of course, the information must still be obtained through information sensing equipment that can sense changes in the state of the object, such as RFID devices, infrared sensors, global positioning systems, laser scanners, temperature and humidity environment sensors, etc. The processing platform through data analysis finally presents the state of the physical world to the user, and realizes the control and monitoring of objects by people.

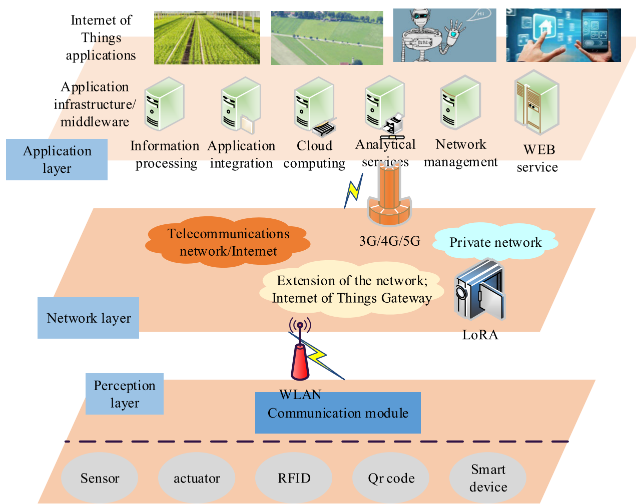


FIGURE 1. IoT technology architecture.

From a functional point of view, IoT can be summarized as the perception, transmission and processing of information between things. Therefore, IoT network architecture is considered to be composed of a perception layer, a network layer, and an application/middleware layer, as shown in Figure 1. Among them, the perception layer mainly implements intelligent perception functions, including information collection, capture, and object recognition. The network layer mainly realizes the transmission and communication of information, and the network layer can rely on the public telecommunication network and the Internet. In the application/middleware layer, the middleware layer mainly implements the interface and capability call between the network layer and IoT application services, including analysis and integration, sharing, intelligent processing, and management of the business. It is embodied as a series of business support platforms, management platforms, information processing platforms, intelligent computing platforms, middleware platforms, etc. The application layer mainly includes various applications, such as monitoring services, smart grids, industrial monitoring, green agriculture, smart homes, environmental monitoring, and public safety.

B. WIRELESS CAMPUS NETWORK

Wireless campus network is a typical application of wireless local area network in campus. Wireless local area network (WLAN) technology gradually matured in the 1990s and put into commercial use. It can not only extend the traditional wired network, but also replace the traditional wired network in some environments [32].

1) CHARACTERISTICS OF WIRELESS LAN

a: SIMPLICITY

The installation of WLAN bridge transmission system is quick and simple, which can greatly reduce the tedious work of laying pipes and wiring.

b: FLEXIBILITY

The wireless technology allows WLAN devices to be flexibly installed and adjusted according to specific locations, so that the wireless network can reach areas where the wired network is difficult or even impossible to cover.

c: LOWER OVERALL COST

The installation of WLAN network can greatly reduce the cost of wiring. In some special environments that require frequent movement and change, WLAN technology can achieve better protection of existing investments. At the same time, because the purpose of WLAN technology itself is a network transmission technology for data communication, it can be directly connected to the intranet of colleges and schools through the 100M adaptive network port.

d: STRONG EXPANDABILITY

WLAN bridge system supports multiple topologies and smooth expansion. It can be easily expanded from a small-capacity transmission system to a medium-capacity transmission system.

2) COMPOSITION OF WIRELESS LOCAL AREA NETWORK

The main components of a wireless local area network are: a wireless network card, a wireless access point (AP), a wireless bridge, an access controller (AC), and an authentication server (AS). It also includes antennas, POE power supply equipment, adapters, etc.

In the current campus network environment, with the help of the wireless AP architecture, a logically independent wireless network can be established on the basis of the existing campus wired network. Since the front-end wireless AP runs, in the normal mode, all wireless data, and control traffic are handled by the wireless controller. Therefore, borrowing the existing campus network switches/routers to form a centralized control and management "overlay" wireless network is a more economical option.

When the AP is connected to the access layer POE switch of the existing wired campus network, the IP address is provided by the wired end of the existing campus network to provide a dynamic address or set a static IP address that meets the IP address planning standards. The wireless network can not only realize the wireless access function of the notebook computer, but also realize the access of the WI-FI mobile phone. Even any network device that supports WI-FI can be connected to the wireless campus network. The wireless campus network is mainly used as the infrastructure part of the wireless video surveillance system in this article. Therefore, the wireless camera can be directly connected to the wireless campus network, and realize the function of wireless video surveillance under the control of the video surveillance server.

C. VIDEO SURVEILLANCE SYSTEM

The video surveillance system [33] can provide very rich and powerful functions. The video processing module embedded in the front end analyzes the monitored picture, and uses intelligent algorithms to compare with the user-defined security model. Once a security threat is discovered, the monitoring center is immediately alerted. The powerful image processing capability of the video processing module and the robust intelligent algorithm can greatly improve the accuracy of the alarm and reduce the probability of false positives and false negatives. In video surveillance systems, the requirements for video processing modules are very high. It not only needs to complete the functions of video encoding and decoding, but also supports the real-time processing of the intelligent video recognition algorithm, so that the system can maximize its advantages.

The monitoring system is generally divided into 3 parts: video monitoring terminal, video monitoring center, and customer browsing terminal, as shown in Figure 2.

Surveillance end: generally refers to tens of thousands of surveillance terminal cameras and related circuits and control equipment of the surveillance system. The monitoring terminal usually has the functions of collecting, encoding, and transmitting integrated video.

Surveillance center: the video surveillance center is the core part of the surveillance system, including access control server, virtual machine, data center, etc. Unlike the background monitoring center of the ordinary monitoring system, the video monitoring center not only realizes the access management of the monitoring and browsing nodes through the control server, but also needs to schedule and run two types of tasks from the monitoring end and the client through a huge virtual machine group. Realize the storage management of massive video data, user demand analysis, system billing, and other functions.

Client: the client browser is the window for users to view and manage the surveillance video. In the video surveillance system, the client browser refers to the video access port that supports various mainstream browsers, usually with user identification, login, video browsing, management, and other functions.

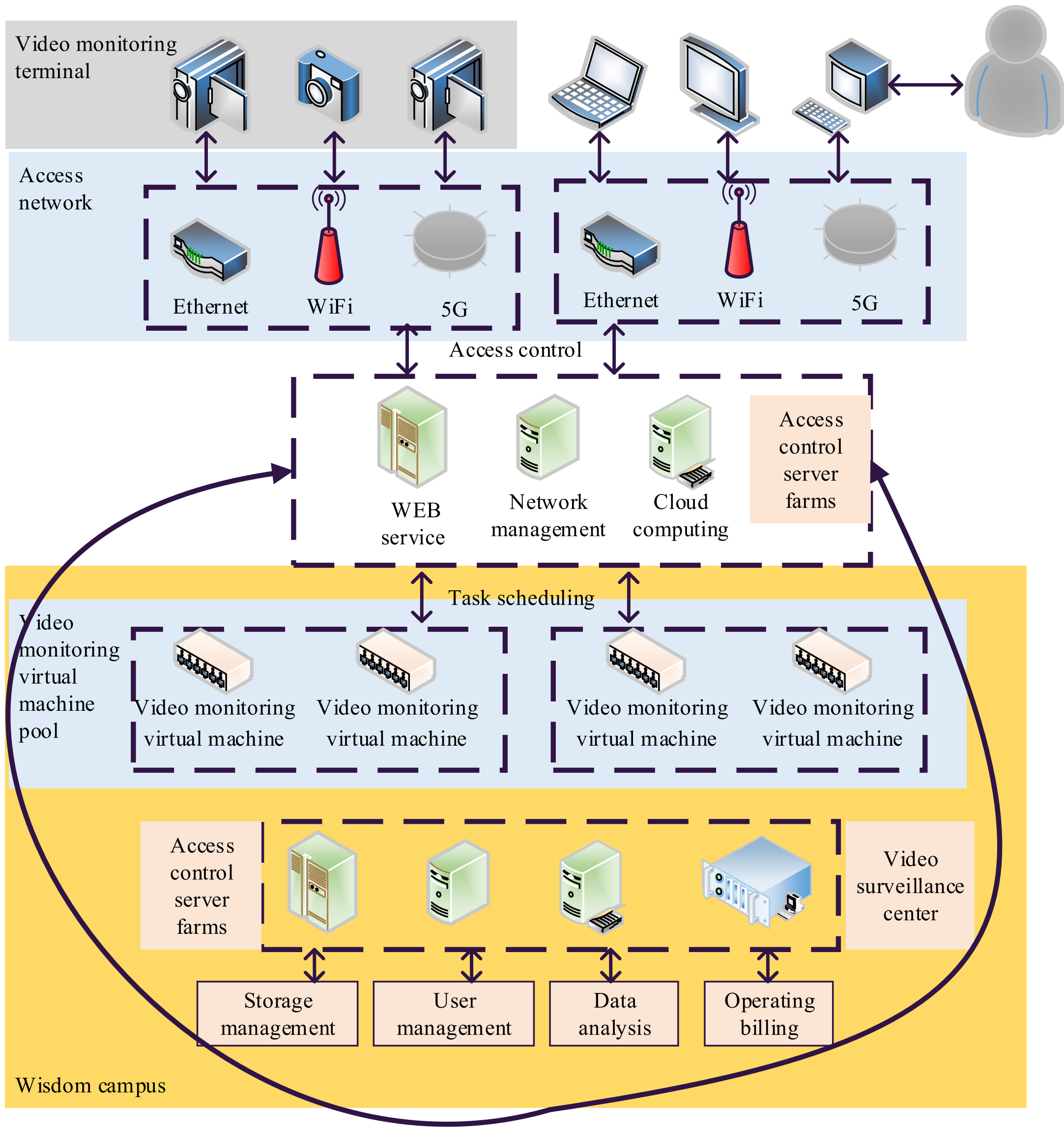


FIGURE 2. Video surveillance system architecture diagram.

Intelligent video surveillance equipment has more powerful image processing capabilities and intelligent factors than ordinary network video surveillance equipment. Therefore, it can provide users with more advanced video analysis functions, which can greatly improve the capabilities of video surveillance systems and enable video resources to play a greater role.

III. OPTIMIZATION OF WIRELESS VIDEO SURVEILLANCE SYSTEM FOR SMART CAMPUS BASED ON INTERNET OF THINGS

A. OVERALL STRUCTURE OF VIDEO SURVEILLANCE SYSTEM

The wireless video surveillance system in the campus is composed of a video acquisition module, a wireless transmission module and the system's general control terminal module. The module's construction mode is used to reduce the construction cost and complexity of the wireless video surveillance system. The use of modular system function construction is helpful to reduce the threat of system collapse caused by the damage of scattered equipment components, reduce the difficulty of building a system platform, and optimize the overall function of the system. The construction of the overall architecture of the wireless video surveillance system is shown in Figure 3. The whole system architecture is divided into three layers: perception layer, network layer and application layer. The perception layer connects the monitoring equipment with sensors and actuators through a wireless network. The network layer is the information transmission medium between the application layer users and the perception layer devices. The application layer is the user's interactive interface to IoT video

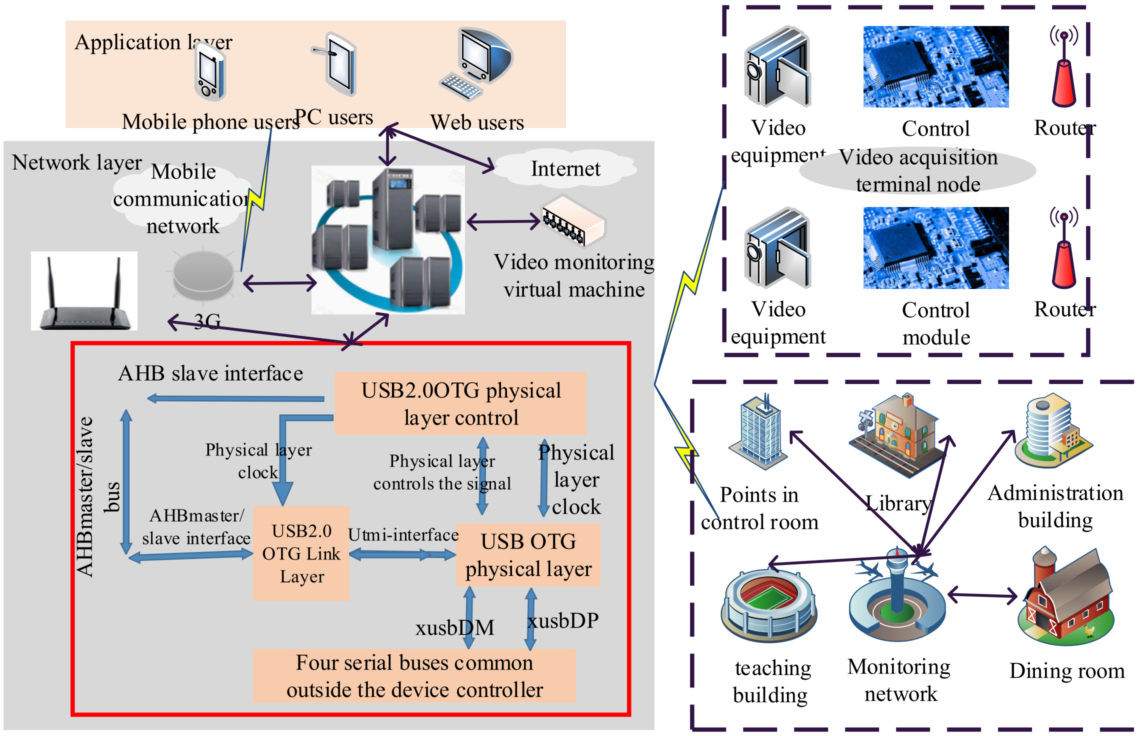


FIGURE 3. The system monitors the overall structure.

surveillance system. As shown in Figure 3, the overall structure mainly includes buildings, sub-monitoring rooms, monitoring centers and monitoring networks. The specific instructions are as follows:

1) Each building mainly includes a teaching building, etc. A number of monitoring points must be deployed at the main entrance and inside of each building, so each building should be equipped with a corresponding node server. In addition, the monitoring point equipment at each major road junction on the campus can be connected to the nearest building.

2) Each building has a sub-monitoring room, which can realize real-time monitoring and basic management of the area.

3) The monitoring center is the highest monitoring management department, which can realize real-time monitoring and comprehensive management of the entire school area.

4) The buildings and other monitoring points are ultimately interconnected through the monitoring network. The monitoring network is designed based on Ethernet. The backbone network can be based on the existing gigabit optical cable in the school, and can also use spare optical cables or be laid separately.

ZigBee's reliability, security, low cost, scalability, and signal coverage are in line with the requirements of smart home control. Using ZigBee technology will enable smart campuses to be wireless at a lower cost. ZigBee is a wireless communication technology with low power consumption, low transmission rate, low cost, low complexity, and short distance. It is especially suitable for places with small amount of data transmission, low power consumption and many network nodes. The intelligent campus is characterized by dense equipment, short sensor communication distance, and small amount of data, which is in line with ZigBee's characteristics. Therefore, this paper adopts ZigBee technology to form a sensor network. Numerous ZigBee nodes are connected with sensors and actuators, and they are responsible for uploading information to the ZigBee coordinator and receiving the command and control campus equipment sent by the coordinator.

B. VIDEO QUALITY OPTIMIZATION

The quality of surveillance video is mainly reflected in performance indicators such as real-time, fluency and clarity. In a general video surveillance system, real-time performance is mainly determined by the video processing performance of the monitoring terminal. Real-time performance requires client software to decode network video data at the fastest rate.

In the remote monitoring system, there is a receiving buffer in the monitoring terminal, which buffers the network data stream, so the real-time nature requires that the data is hardly buffered in the buffer. The smoothness of the video is mainly determined by the relative relationship between the decoding and display rate of the monitoring terminal and the receiving rate of the network video stream. When the decoding and display rate is higher than the video stream receiving rate, the decoding and display will block waiting for the video to receive. Since this waiting process is not evenly distributed in every frame, it will cause the problem of fluency in playing video. The resolution is mainly related to the packet loss rate of the video data. Under ideal network conditions, the packet loss rate is mainly related to the data processing efficiency of the monitoring terminal. When the terminal data processing rate is low, due to the accumulation of video data, it will cause the receive buffer to overflow, which will lead to an increase in the packet loss rate and cause clarity problems. In practical applications, network congestion greatly affects the packet loss rate, which determines the clarity of the video.

It can be seen from the above analysis that in order to ensure the real-time performance of the video, the processing performance of the video terminal must be improved. Due to the real-time requirement, the monitoring terminal needs to process data at the fastest rate, and the fluency requires the video stream to be buffered, and at the same time, the decoding and display rate of the monitoring terminal is controlled. In order to achieve a good monitoring effect, a balanced control strategy must be found. On the one hand, let the video buffer to ensure that the video decoding and display are not blocked. On the other hand, it guarantees that the data is quickly decoded without cumulative delay. At the same time, network congestion must be suppressed to ensure video clarity.

1) ZERO COPY BUFFER STRATEGY

The amount of data processed in streaming media programming is very large. Reducing data copying can increase the processing speed of the client's streaming media data, reduce latency, and reduce processor load. At the same time, it can also reduce the packet loss caused by the client due to data copying, too late to process subsequent data packets. Thereby saving system resources and improving the playback quality of streaming media. The zero-copy buffer strategy can reduce data copy work by 90% through a reasonable buffer design and greatly improve system performance. The zero-copy buffer strategy combines the receiving buffer and the decoder input buffers to make the data copy operation between the buffers become a pointer operation.

The schematic diagram of the zero copy buffer is shown in Figure 4. Read_ptr represents the video decoding pointer, which points to the data to be decoded. Write_ptr represents the received data pointer, pointing to the storage address of the network data. Valid_data_ptr represents the first address of the effective buffer, and the first address where network data is stored. The backup buffer does not store the video stream received from the network. When the valid data is divided into 2 blocks, and the decoded data exists in 2 parts of the buffer, the pass to the decoder Read_ptr cannot meet the requirements. Therefore, the backup buffer needs to be used. The specific strategy is to copy the data after Read_ptr to Buffer_ptr, so that the decoded data becomes a continuous buffer. Because in the video surveillance system, the data volume of a frame of data is much smaller than the receive buffer. Therefore, the probability of such copying is very small, and the amount of data copied each time is also very small, which can greatly optimize the performance of the system and improve the efficiency of decoding and display of the monitoring terminal.

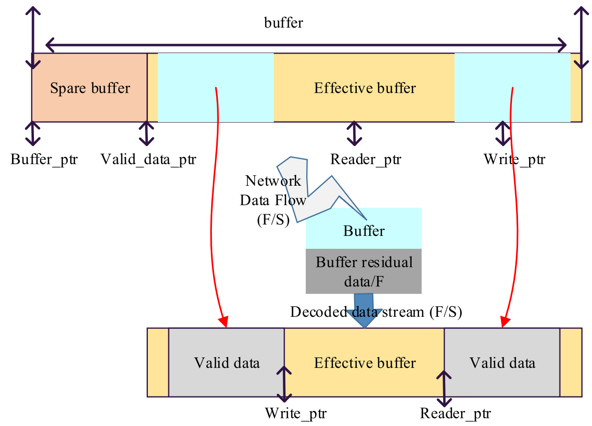


FIGURE 4. Buffer principle and state diagram.

2) NETWORK CONGESTION SUPPRESSION STRATEGY

When there are too many packets to be transmitted in the communication network, the performance of the entire network will be reduced, the transmission quality will be reduced, and network congestion will occur. When network congestion occurs, if the network congestion cannot be suppressed in time, the video delay will rise and the network packet loss rate will increase sharply. At the same time, it will also bring about certain fluency problems, which will have a great impact on the video quality. When network congestion occurs, the rate of server-side video acquisition is appropriately reduced, which not only reduces the transmitted data, reduces the network load, but also reduces the client's data requirements, reducing the chance of video quality degradation.

The client can periodically count the total number of received data packets and the number of lost data packets. Then, fill the data packet according to the data packet format of RTCP and send it to the server. Finally, the server side can dynamically adjust the data packet collection and transmission rate through the corresponding flow control algorithm and the specific parameters transmitted from the client.

3) CODEC RATE COORDINATION STRATEGY

In general video surveillance systems, when monitoring terminals perform video decoding, in order to ensure the integrity of each frame of data, it is necessary to determine whether the data in the receive buffer reaches a certain limit Limit_A, but because the monitoring image is fixed in the background and the background is in vigorous motion, each frame the amount of data differs greatly, and the number during intense exercise is often several times that at rest, so the choice of Limit_A is more difficult. When Limit_A is selected to be small, in the case of violent motion, the decoded data may not be a complete frame, causing video quality problems. When Limit_A is selected to be large, it may cause video pauses and long video delays under almost static conditions. Therefore, Limit_A must be dynamically changed, and at the same time, the data in the buffer area is different due to the amount of data at rest and in motion. The amount of data must also be strictly controlled to prevent large-scale delays at rest.

C. MULTI-CAMERA TRACKING OPTIMIZATION

Different feature descriptors have different scenarios and effects. To apply the description operator to target tracking, whether it is the classic SUFT operator [34] or the newly proposed BRISK operator [35], ORB operator [36], etc., the detection of the entire picture is time-consuming long. It is unrealistic to use detection operators on the entire image. Therefore, we naturally apply the detection operator to the foreground area of the target, that is, to effectively extract the features of the target on a small scale. The insensitivity of SIFT operator to deformation, illumination and scale provides a good basis for its wide application. Therefore, this paper chooses SIFT operator.

Since each moving target in the monitoring system may be tracked by multiple cameras at the same time, automatically selecting the camera with the best position to track the target plays an important role in improving monitoring performance and effectively allocating computing resources. In order to effectively track multiple cameras collaboratively between moving targets, the system introduces the tracking optimization algorithm POSRCA.

There are usually important moving targets that need priority tracking in the monitoring system. The tracking optimization algorithm POSRCA performs data fusion according to the tracking priority of the target and the occlusion state of the target in each camera and the size of the segmented image. The weight function W is defined as follows:

$$W_{i,j} = O_{i,j}(aP_j + bS_{i,j}) \tag{1}$$

In the formula, the variable $W_{i,j}$ represents the weight of the target j in the field of views of the camera i. The variables $O_{i,j}$ represent the occlusion state of the target j in the field of view of the camera i. The variable P_j represents the tracking priority manually assigned to the target j. The variables $S_{i,j}$ represent the divided image size of the target j in the field of view of the camera i. The variables a and b represent the weight parameters of the target tracking priority and the size of the segmented image in the calculation of the weight function, and are dynamically adjusted according to the highest priority and the resolution of the video frame. If the target j in the view field i of the camera is in the occlusion state, the variable $O_{i,j}$ is 0.1, otherwise it is 1. The detailed steps of the tracking optimization algorithm POSRCA are as follows:

1) For the target j that needs to be assigned a camera, the central server calculates the weight function value $W_{i,j}(i \in Q)$ of the target in the field of view of camera i, Q represents the set of all cameras that can see the target j.

2) Find the maximum weight function value $W_{m,j} = \max_{i \in Q} W_{i,j}$ of target j, assign target j to camera m for tracking, and update the camera tracking information stored in the central server database.

3) If there are N targets being tracked by camera m, in order to dynamically balance the tracking load and computing resources of the camera, the lowest priority tracking target in camera m is redistributed among the remaining cameras according to the above steps.

After assigning the best camera to the target for tracking, the remaining cameras that can see the target will not track it. The following situations will trigger the central server to execute the above optimization algorithm on a target:

1) Automatically refresh the weight function value of each tracked target in all visible cameras every certain time, and reassign the best camera to all targets.

2) The new target enters the surveillance field of view of the camera.

3) The target disappears from the assigned camera field of view.

4) The target appears blocked in the assigned camera.

5) Manually start the optimization algorithm to reassign cameras to a certain target.

In order to measure the tracking effect of the tracking optimization algorithm POSRCA in the entire monitoring system in real time, the system introduced a function POT to quantitatively describe the performance of the tracking system. Define the function POT as:

$$\begin{aligned} P_{all} &= POT(C_1, C_2, \ldots, C_r; Q_1, Q_2, \ldots, Q_m) \\ &= \sum_{i=1}^{n} \sum_{j \in \Theta_i} POT(C_i; Q_j) = \sum_{i=1}^{n} \sum_{j \in \Theta_i} W_{i,j} \end{aligned} \tag{2}$$

In the formula, the variable P_{all} represents the POT of the entire monitoring system. The variable Θ_i represents the set of all targets assigned to the camera i for tracking. The variable $POT(C_i; Q_j)$ represents the POT assigned by the camera i to the target j. The variable $POT(C_i; Q_j)$ assigned to the system is equal to the weight function value $W_{i,j}$ of the target j in the field of view of the camera i. Therefore, the higher the priority of the target tracked by the camera, the larger the segmented image and the non-occluded state, the higher the POT assigned by the camera to the target and the better the tracking effect.

D. OPTIMIZATION OF MONITORING EVENTS BASED ON BAYESIAN MODEL

Surveillance Event Detection (SED) is the core part of the video surveillance system and a key part that determines the overall accuracy of the system and the detection experience. Only when the system can accurately determine the monitoring event can it be able to carry out early warning and alarm for the monitoring event. There are many ways to discriminate between monitoring events, and action recognition can be considered as a preliminary study of monitoring event detection. Bayes' rule describes the method of finding the probability of occurrence of an event of interest B under known conditions. The mathematical expression is as follows:

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)} \alpha L(B|A)\Pr(A) \tag{3}$$

In an actual video surveillance system, we can understand the events of interest A as the surveillance events we need to detect, and the known conditions B as the underlying features that we can observe. Based on a large number of statistics, we can get more accurate $\Pr(A|B)$. This number can largely help the system to predict the events that will happen by knowing the underlying features. However, in order to arrive at an accurate prior probability $\Pr(A|B)$, it often requires a large amount of statistical data, which is inefficient in actual execution. Therefore, using Bayes' rule can transform the problem of calculating the prior probability into the problem of obtaining the posterior probability $\Pr(B|A)$. Since the number of occurrences of A events will be much less than the number of occurrences of B observations, it is relatively easy to obtain the posterior probability, so that the required prior probability can be calculated more easily.

Bayes' rule is based on a single input and a single output. In practical applications, there are often multiple inputs. Consider that there are multiple underlying features $F = \{f_1, f_2, \ldots, f_n\}$, the solved event is I, and the underlying feature is not directly connected to the solved event, but has other intermediate variables. The intermediate variable is recorded as $R = \{r_1, r_2, \ldots, r_m\}$. Then we can get that the relationship between the underlying feature F and the intermediate variable r_m is expressed by Bayes' rule, there is

$$p(r_M|F) = \frac{p(f_1f_2f_3\ldots f_n|r_M) * p(r_M)}{p(f_1f_2f_3\ldots f_n)} = \frac{p(f_1f_2f_3\ldots f_n r_M)}{p(r_M)} \tag{4}$$

Assuming that the underlying features are independent of each other, using chain rule on the molecule of formula (4), there is

$$\begin{aligned} p(r_M|F) &= p(f_1|f_2\ldots f_n r_M)p(f_2|f_3\ldots f_n r_M)p(f_s\ldots f_n r_M) \\ &= p(r_M)\prod_i p(f_i|r_M) \end{aligned} \tag{5}$$

Similarly, for the denominator, we have

$$p(f_1f_2f_3\ldots f_n) = \sum_j p(r_j)\prod_i p(f_i|r_j) \tag{6}$$

Therefore, the relationship between the derivation of the underlying feature F and the intermediate variable rM can be expressed by Bayes' rule as

$$p(r_M|F) = \frac{p(r_M)\prod_i p(f_i|r_M)}{\sum_j p(r_j)\prod_i p(f_i|r_j)} \tag{7}$$

Similarly, the relationship between the underlying features F and events I can be expressed by Bayes' rule as

$$p(I|F) = \frac{p(F|I)p(I)}{p(F)} = \frac{p(f_1f_2f_3\ldots f_n I)}{p(f_1f_2f_3\ldots f_n)} \tag{8}$$

Applying the chain rule to the molecule in formula (8), there is

$$\begin{aligned} p(f_1f_2f_3\ldots f_n I) = p(I = i_K)\sum_{R_1\ldots R_j}[\prod_i p(f_i|r_1 = R_1,\\ \ldots, r_j = R_j)\prod_j p(r_j = R_j|I = i_K)] \end{aligned} \tag{9}$$

Similarly, applying the chain rule to the denominator in formula (8), there is

$$\begin{aligned} p(f_1f_2f_3\ldots f_n) = \sum_{R_1\ldots R_jk} p(I = i_K)\prod_i p(f_i|r_1 = R_1,\\ \ldots, r_j = R_j)\prod_j p(r_j = R_j|I = i_K) \end{aligned} \tag{10}$$

The required probability can be calculated by substituting formula (9) and formula (10) to formula (8).

In this paper, the introduction of Bayesian-based cognitive processes can reduce the need for high-level picture features based on complex technologies, thereby reducing computational complexity, computing time, increasing algorithm robustness, and reducing system modeling complexity degree.

This paper builds a Bayesian with a three-layer hierarchical structure, which is composed of a feature layer, a derivation layer, and a semantic layer. Using the feature layer as the bottom layer, the bottom-up derivation is performed to achieve the purpose of deducing the high-level semantics from the bottom-layer features. Therefore, we propose a three-layer hierarchical structure based on Bayes as shown in Figure 5.

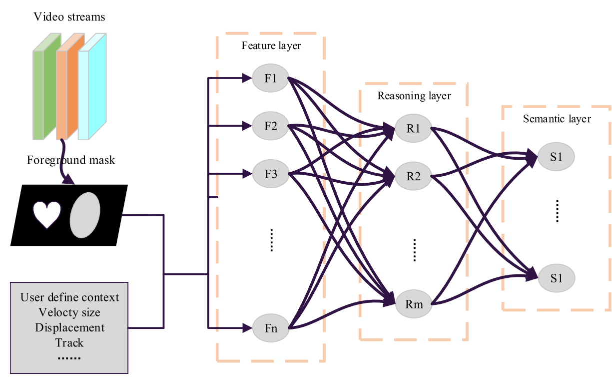


FIGURE 5. **Bayesian three-layered structure.**

The addition of the derivation layer strengthens the derivation relationship between the information of the feature layer and the events of the semantic layer, and makes the combination of diverse information of the feature layer. It is possible to derive semantic layer events from these combined forms again. At the same time, it is possible to enhance the impact of individual feature layer information on semantic layer events. Therefore, the internal relations of the Bayesian system as a whole are greatly improved simultaneously. The introduction of the derivation layer greatly reduces the requirement for high-level feature extraction. In this way, the introduction of hierarchical structure reduces the complexity of feature extraction based on Bayesian monitoring event detection, thereby reducing the time required for the overall operation of the system. In addition, the addition of the derivation layer diversifies the relationship between feature layer information and semantic layer events, which is conducive to finding a better solution. By setting appropriate parameters and network structure, it is possible to detect the monitoring events with high accuracy and strong robustness.

We can deduce the semantic layer events through the status of the feature layer information. Remember that the feature layer elements are $\{F_1, F_2, \ldots, F_n\}$, the derivation layer elements are $\{R_1, R_2, \ldots, R_m\}$, and the semantic layer events to be derived are S. It can be obtained from the Bayesian network calculation rule. From the feature layer elements $\{F_1, F_2, \ldots, F_n\}$, the probability that the semantic layer event S is in a certain state is calculated by formula (11):

$$\begin{aligned} & p(S = s_o|F_1 \ldots F_n) \\ & = \frac{p(F_1 \ldots F_n|S = s_o)p(S = s_o)}{p(F_1 \ldots F_n)} \\ & = \frac{\sum p(S = s_o) \prod p(F_i = f_i^q|R_i = r_g) \prod p(R_j = r_j|S = s_o)}{\sum p(S = s_k) \prod p(F_i = f_i^q|R_i = r_g) \prod p(R_j = r_j|S = s_k)} \end{aligned} \tag{11}$$

Among them, the variable $p(R_j = r_j|S = s_o)$ represents the probability that the derivation layer semantics R_j takes the value r_j when the semantic layer events take the value s_o. This probability is given by the probability dependence of the semantic layer events and the derivation layer semantics, that is, the relationship matrix $a^S_{R_\partial}$ between the two. The variable $p(F_i = f_i^q|R_i = r_g)$ represents the probability that the feature layer information F_i takes a value f_i^q of when the derivation layer semantic set related to the feature layer information F_i takes a certain value combination r_g. The probability is given by the probability dependence of the semantic combination of the derivation layer R_i and the feature layer information F_i, that is, its relationship matrix $a^{R_i}_{F_i}$. The variable $p(S = s_k)$ represents the probability when the semantic layer event S takes the value s_k. This probability is initially given by statistical data. For the relationship matrix a^A_B between the system element A and the system element B, that is, the probability dependence of the probability that the system element B is in a certain state is derived from the state of the system element A, which can be obtained through statistical data, that is, expert knowledge. The size of the relationship matrix is determined by the size of the system elements.

Remember that the system element A has N_A possible states $\{a_1, \ldots, a_{N_A}\}$, and the system element B has N_B possible states $\{b_1, \ldots, b_{N_B}\}$. The state is, then the relationship matrix a^A_B size is $N_A \times N_B$. For the feature layer information F_i, there may be a series of derivation layer semantics and if it is related, the semantic set $R_i = R_i'$ of these derivation layers is recorded, and the size of the relationship matrix $a^{R_i}_{F_i} = a^{R_i\prime}_{F_i}$ is $N_{F_i} \times N_{R_i\prime} = N_{F_i} \times \prod R_j$. Among them, the matrix $a^{R_i\prime}_{F_i}(u, v)$ element represents the probability that the feature layer information F_i takes the vth state when the set of semantics of the derivation layer is R_i', and the uth specific combination of states is taken. By customizing the combination of different matrix elements and substituting formula (9) for calculation, it can affect the final monitoring event detection effect.

Due to the large number of parameters to be determined in BN Bayesian model and the huge search space, the EM algorithm is often used to conduct optimization under strict requirements on the selection of initial conditions, otherwise the obtained optimization results will be relatively uneven, and it is difficult to obtain the global optimal solution. Moreover, because the algorithm idea of EM algorithm is to adjust the parameters of the target model and the target algorithm, the structure adjustment ability of the model is relatively low, and it can only play a limited optimization role in Large Scale models with insufficient expert knowledge. Therefore, considering the characteristics of BN Bayesian network model, such as the high dimension of adjustment parameters and the need for structural adjustment, we consider to adopt a genetic algorithm more suitable for large-scale parameter adjustment.

In order to further improve the robustness and detection accuracy of the system, this paper uses a genetic algorithm to optimize and adjust the overall system parameters and structure on the Bayesian network established on the basis of expert knowledge. Genetic algorithm is a highly parallel, random, and adaptive global optimization algorithm based on "survival of the fittest". It expresses the solution of the problem as the survival process of the "chromosome" of the fittest, through continuous iteration of the "chromosome" group, including operations such as reproduction, crossover, and mutation, until a certain performance is met the indicators and convergence conditions are terminated, so as to find the optimal solution to the problem.

Considering the specific characteristics of the Bayesian network, that is, strong specificity, the complicated mathematical relationship between the parameters, the large dimension of the parameter space, and the need for quantifiable evaluation, etc., we chose the real number coding method to model the Bayesian network GA. Bayesian-oriented GA modeling is shown in Figure 6.

GA believes that every possible BN model is an individual in a population. Considering a BN model with N + 1 nodes (including feature layer, derivation layer, and semantic layer), there are a total of N relational matrices. Convert each relational matrix into a Gene in the individual, and there are n genes in total. Each Gene is divided into three parts, namely, the Gene Index of P1, the starting position (S) and ending position (E) of the corresponding relational matrix of P2 (directed arrow), and the element of the relation matrix of P3. The elements of the relational matrix are mapped to the one-dimensional sequence in the order of the first and last columns of the two-dimensional matrix, and arranged in the last part of the gene. Since the number of BN nodes related to each Gene is different, a Flag of Gene Start is needed at the front end of the sequence of relational matrix elements

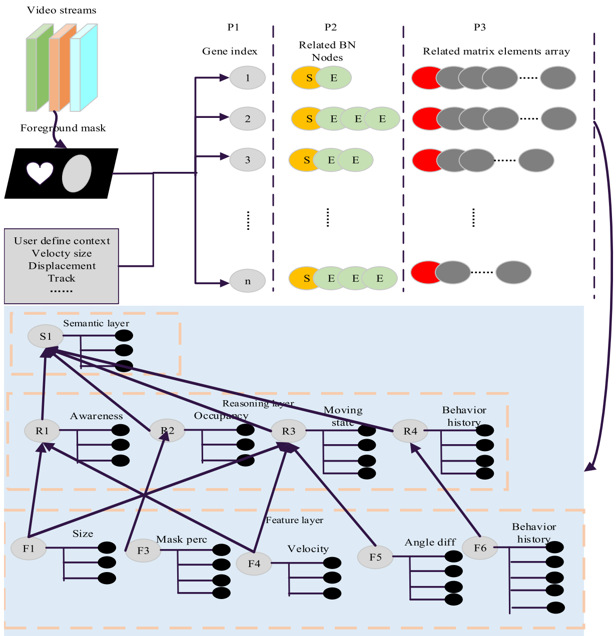


FIGURE 6. GA real coding model for Bayes.

to mark the starting position of the arrangement of relational matrix elements.

IV. SIMULATION EXPERIMENT

A. VIDEO QUALITY OPTIMIZATION

The test content of this experiment is mainly for the system state when the client sends a connection request 5 s, 10 s, 30 s, 60 s, 90 s, 120 s, 150 s, 200 s, 250 s, 300 s. These system states mainly include delay, packet loss rate, and the percentage rate of frames that generate pauses. When the test process is single user, compare the video quality before and after optimization. As shown in Figure 7(a), the delay changes before and after optimization. It can be observed from Figure 7(a) that the delay before optimization increases with the increase of the monitoring time. This is due to the excessive copy operations between the monitoring terminal buffers, which causes the client to decode the display and the speed is relatively low, thus causing the cumulative delay of the monitoring terminal. The optimized delay is basically stable, about 1.5 s, and there is no cumulative delay. This is due to the control of the buffer strategy, and the video decoding and display speed of the monitoring terminal has been greatly improved. At the same time, under the control of the codec rate coordination strategy, the video codec speed is relatively average, which effectively suppresses the cumulative delay and ensures the real-time performance of the system. Therefore, the optimized control strategy is very effective for delay control.

As shown in Figure 7(b), it is a graph of the change of packet loss rate before and after optimization. It can be observed from Figure 7(b) that before optimization, the packet loss rate of the system is relatively stable at the beginning of monitoring. But with the increase of monitoring time, the packet loss rate increases rapidly. Due to the existence of accumulated delay, the buffer must overflow. Therefore, the packet loss rate increases rapidly. In the optimized system, since there is no cumulative delay, the system buffer utilization rate is relatively small, which will not cause packet loss due to buffer overflow. Therefore, the monitoring video quality of the system is ensured. In the optimized system, the packet loss rate still fluctuates greatly. This is because in the case of network congestion, a large packet loss rate is inevitably generated. However, due to congestion control, the packet loss rate will be suppressed and slowly return to normal levels.

Figure 7(c) is a graph of the percentage change of paused frames. As can be seen from Figure 7(c), the percentage of paused frames after optimization has been greatly improved compared with that before optimization. The percentage of paused frames after optimization is probably stable at about 1.5%. With the improvement of real-time video, the time of video buffering is also greatly reduced, and the amount of data in the buffer is also greatly reduced. As a result, the monitoring terminal decodes and threads intermittently wait for the network data stream, which results in an increase in the pause percentage and pauses in the video. Since the optimized video pause percentage is basically stable and in an acceptable range, it further indicates that the codec rate coordination strategy is effectively controlled.

B. MULTI-CAMERA TRACKING OPTIMIZATION

BRISK algorithm mainly USES FAST9-16 to detect feature points. It has good rotation invariance, scale invariance, and robustness. In the ORB scheme, FAST is first used as the feature point detection operator, and then the improved BRIEF is used for descriptor calculation. SURF algorithm USES fast Hessian algorithm to detect key points, and SURF operator describes the surrounding area of key points through an eigenvector. SIFT features are based on a few points of interest in the local appearance of an object irrespective of the size and rotation of the image. SIFT features have a large amount of information, suitable for rapid and accurate matching in massive databases.

Figure 8 and Figure 9 are the statistical results of the four descriptors. It can be seen from Figure 8 and Figure 9 that although SIFT has higher complexity than several other algorithms, it has relatively good detection and matching effects on a small scale. Other algorithms are relatively small in the number of feature points and the number of matches. Too few matching points may have a greater impact on the subsequent judgment of target accuracy, resulting in a reduction in the robustness of the judgment results. Although the SIFT descriptor is higher in calculation time and complexity, in small-scale application scenarios, the SIFT descriptor can effectively extract the key points of the target, and the number of target key points and the number of matching key points are more common than other commonly used calculations. The child has improved significantly. Therefore, this paper chooses the SIFT descriptor.

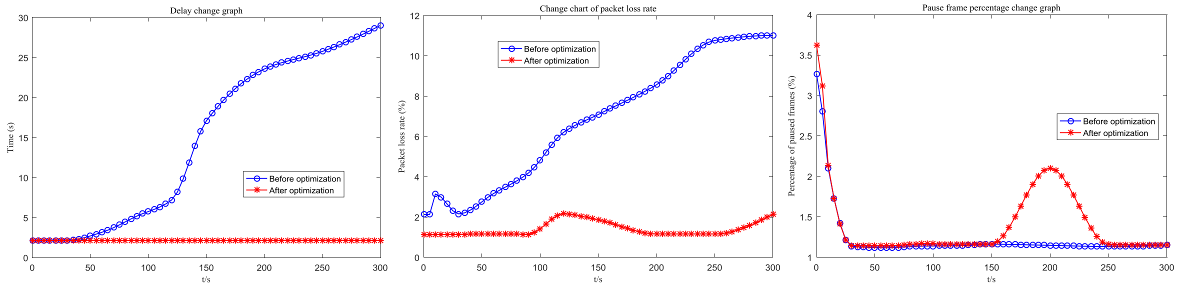
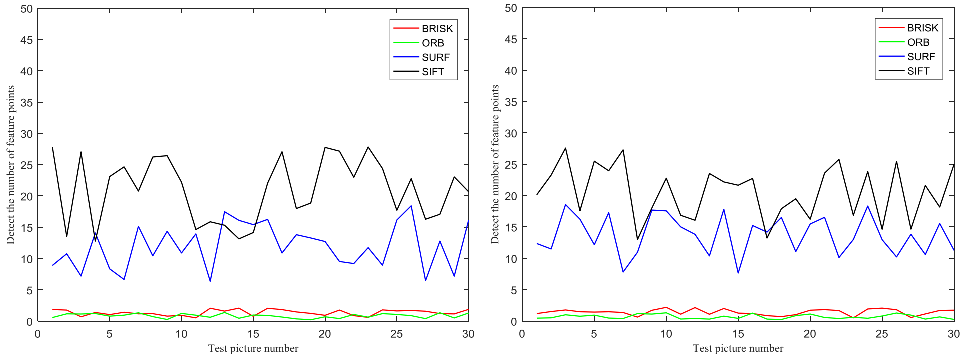


FIGURE 7. Changes in delay, packet loss rate, and percentage of paused frames before and after optimization.



(a) The 71-th frame (b) The 72-th frame

FIGURE 8. Number of detected feature points in two adjacent frames.

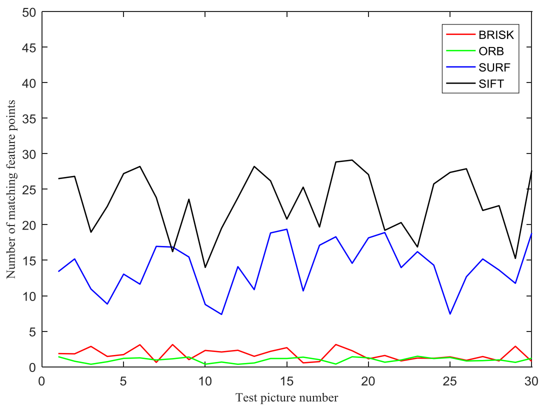
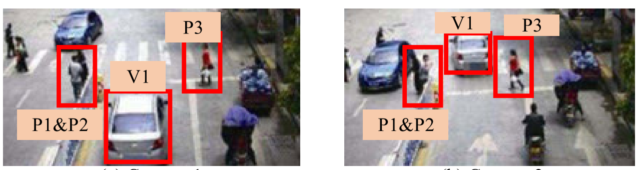


FIGURE 9. Number of matching feature points.



(a) Camera 1 (b) Camera 2

FIGURE 10. Target matching results between cameras in outdoor scenes.

Figure 10 shows the surveillance images of two cameras in the outdoor scene at the 10th s. The priority assigned to all targets in the experiment is 1. Since the camera in this outdoor scene is far away from the monitoring ground plane and the target, the central server performs target consistency matching through the mapping between multiple cameras, and concludes that Person1, Person2, Person3, and Vehicle1 are in the overlapping area of the two cameras. Because Person1 and Person2, Vehicle1 and Person3 block each other in Camera1, after the fusion calculation of the algorithm POSRCA, Camera2 is assigned to track all targets at this moment.

Figure 11 shows a statistical chart of the results of the number of objects in the region of interest and the coverage of the visual field. Among them, T_i represents the target of interest. As can be seen from the above figure, when the number of interested objects in the system increases rapidly, the coverage of the system's field of vision will decrease, but it basically remains at a high level. As the Nash equilibrium is reached, the cameras in the system will gradually cover the entire area. At the same time, the resolution of the field of view of the target of interest has reached a fairly high level. So far, we have verified the effectiveness of the algorithm in this paper.

In order to better evaluate the tracking performance achieved by the fusion algorithm POSRCA, the system introduced the occlusion switching algorithm (OSA) [37] to compare the tracking performance between them. Figure 12(a) shows the tracking performance results obtained when there is only Person4 in the entire monitoring scenario, and the tracking priority is set to 2. Since Person4 will not be blocked in the experiment, only when Person4 leaves the field of view of a camera, the algorithm OSA will switch to other cameras for tracking. When Person4 enters the overlapping area of Camera1 and Camera2, the algorithm OSA will keep the

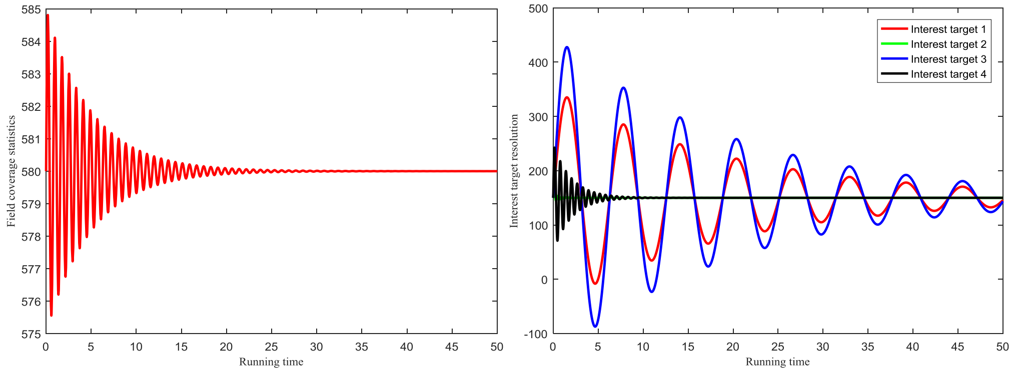
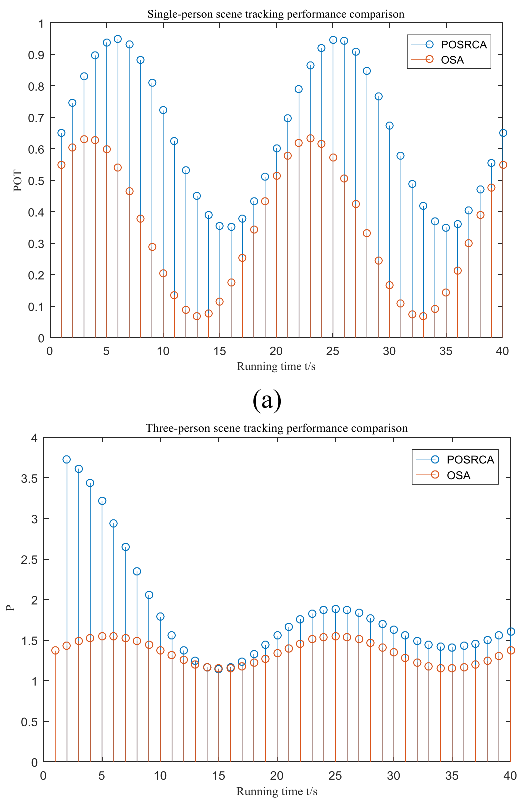


FIGURE 11. Result graph of the number of targets in the region of interest and the coverage of the visual field.



(a)

(b)

FIGURE 12. Track performance evaluation results.

original camera Camera1 to track it. The algorithm POSRCA will automatically switch to Camera2 with a higher weight to track it. Figure 12(b) shows the total tracking performance achieved by the three individuals. It can be seen from the figure that the algorithm POSRCA achieves better tracking performance than the algorithm OSA.

C. BAYESIAN-BASED MONITORING EVENT DETECTION SCHEME

Due to the large differences in the structure of different individuals after GA optimization, the detection accuracy of the individual fluctuates within a small range. Therefore, in the experiment, we integrate the results after 10 GA optimizations as a reference.

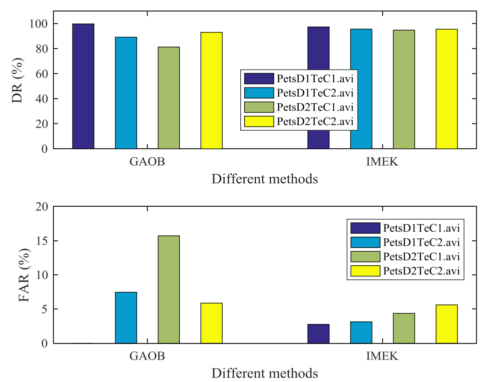


FIGURE 13. Comparison of results based on GA and Bayesian modeling methods based on expert knowledge.

In order to reflect the accuracy and robustness improvement that GA brings to the BN-based system during the optimization process, Figure 13 shows the initial modeling based on expert knowledge (denoted as IMEK) and the model after GA optimization (recorded For GAOB). We can see that in the worst case, GAOB can still maintain a DR close to 95%. The IMEK dropped rapidly to about 81%. This means that the model after GA optimization has stronger robustness and detection accuracy, which is attributed to GA's optimization of the structure and parameters of the model, so that it adjusts and corrects the incomplete parts of expert knowledge.

Figure 14 shows four examples of detected objects. Figure 14(a) shows that the vehicle enters the ROI by reversing and then drives away. Figure 14(b) shows the bicycle passing by the ROI. Figure 14(c) shows a group of people passing by the ROI. Figure 14(d) shows a person passing by the ROI. We compare the GAOB method with the IMEK method and the real state value curve (denoted as GT). It can be seen that after the GA adjustment, the output curve of the system is smoother, the transition is smoother, and the variance is smaller.

The quality of surveillance video is mainly reflected in performance indexes such as real-time, fluency and clarity.

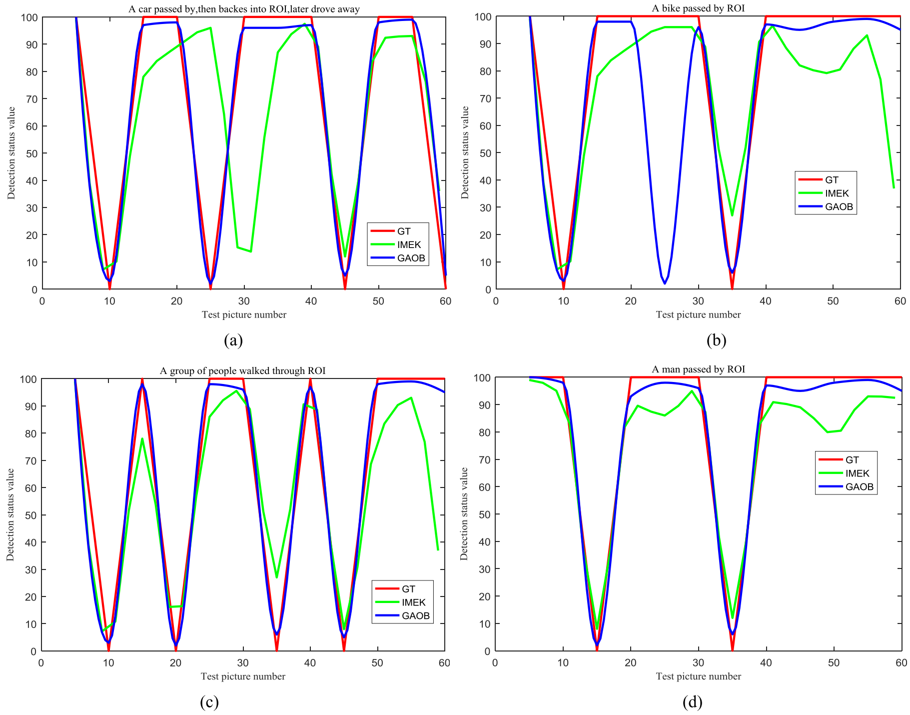


FIGURE 14. State curve based on Bayesian modeling scheme.

Simulation results show that the proposed method can achieve better performance in these three aspects.

V. CONCLUSION

This article first optimizes the surveillance quality in the video surveillance system, effectively improves the quality of the surveillance video, and ensures the real-time, fluency, and high-definition of the surveillance video. Secondly, for the distributed wide-area video surveillance system, a tracking optimization method based on multi-camera fusion is proposed to achieve efficient target consistency matching between multiple cameras. Finally, this paper constructs a Bayesian monitoring event modeling method based on genetic algorithm, which can reduce system redundancy, reduce optimization complexity, increase system modeling flexibility, and reduce the amount of data required for optimization. Simulation experiments show that the optimized video surveillance system can effectively realize the cooperative tracking of multiple cameras in a wide-area surveillance scene, achieve high tracking and surveillance performance, and meet the requirements of smart campus construction.

REFERENCES

[1] C.-T. Yang, S.-T. Chen, J.-C. Liu, R.-H. Liu, and C.-L. Chang, "On construction of an energy monitoring service using big data technology for the smart campus," *Cluster Comput.*, vol. 23, no. 1, pp. 265–288, Mar. 2020.

[2] I. Yaqoob, I. A. T. Hashem, A. Ahmed, S. M. A. Kazmi, and C. S. Hong, "Internet of Things forensics: Recent advances, taxonomy, requirements, and open challenges," *Future Gener. Comput. Syst.*, vol. 92, pp. 265–275, Mar. 2019.

[3] Y. Lin, J. Yang, Z. Lv, W. Wei, and H. Song, "A self-assessment stereo capture model applicable to the Internet of Things," *Sensors*, vol. 15, no. 8, pp. 20925–20944, Aug. 2015.

[4] J. Li, G. Feng, W. Wei, C. Luo, L. Cheng, H. Wang, H. Song, and Z. Ming, "PSOTrack: A RFID-based system for random moving objects tracking in unconstrained indoor environment," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4632–4641, Dec. 2018.

[5] W. Wei, M. Guizani, S. H. Ahmed, and C. Zhu, "Guest editorial: Special section on integration of big data and artificial intelligence for Internet of Things," *IEEE Trans. Ind. Informat.*, vol. 16, no. 4, pp. 2562–2565, Apr. 2020.

[6] M. Noura, M. Atiquzzaman, and M. Gaedke, "Interoperability in Internet of Things: Taxonomies and open challenges," *Mobile Netw. Appl.*, vol. 24, no. 3, pp. 796–809, Jul. 2019.

[7] M. Chen, S. Lu, and Q. Liu, "Uniform regularity for a Keller–Segel–Navier–Stokes system," *Appl. Math. Lett.*, vol. 107, Sep. 2020, Art. no. 106476.

[8] W. Wang, F. Xia, H. Nie, Z. Chen, Z. Gong, X. Kong, and W. Wei, "Vehicle trajectory clustering based on dynamic representation learning of Internet of vehicles," *IEEE Trans. Intell. Transp. Syst.*, early access, Jun. 12, 2020, doi: 10.1109/TITS.2020.2995856.

[9] P. K. Atrey, B. Trehan, and M. K. Saini, "Watch me from distance (WMD): A privacy-preserving long-distance video surveillance system," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 15, no. 2, pp. 1–18, Jun. 2019.

[10] N. Eswara, S. Ashique, A. Panchbhai, S. Chakraborty, H. P. Sethuram, K. Kuchi, A. Kumar, and S. S. Channappayya, "Streaming video QoE modeling and prediction: A long short-term memory approach," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 3, pp. 661–673, Mar. 2020.

[11] M. B. Yahia, Y. L. Louedec, G. Simon, L. Nuaymi, and X. Corbillon, "HTTP/2-based frame discarding for low-latency adaptive video streaming," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 15, no. 1, pp. 1–23, Feb. 2019.

[12] R. Muzaffar, E. Yanmaz, C. Raffelsberger, C. Bettstetter, and A. Cavallaro, "Live multicast video streaming from drones: An experimental study," *Auto. Robots*, vol. 44, no. 1, pp. 75–91, Apr. 2020.

[13] S. Pu, "Development and application of sports video analysis platform in sports training," *J. Adv. Comput. Intell. Intell. Informat.*, vol. 23, no. 1, pp. 139–145, Jan. 2019.

[14] S. Kesavan and J. Jayakumar, "Effective client-driven three-level rate adaptation (TLRA) approach for adaptive HTTP streaming," *Multimedia Tools Appl.*, vol. 77, no. 7, pp. 8081–8114, May 2018.

[15] Y. Li, L. Chu, Y. Zhang, C. Guo, Z. Fu, and J. Gao, "Intelligent transportation video tracking technology based on computer and image processing technology," *J. Intell. Fuzzy Syst.*, vol. 37, no. 3, pp. 3347–3356, Oct. 2019.

[16] Q. Guo, W. Feng, C. Zhou, C.-M. Pun, and B. Wu, "Structure-regularized compressive tracking with online data-driven sampling," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 5692–5705, Dec. 2017.

[17] J. Dhont, J. Vandemeulebroucke, D. Cusumano, L. Boldrini, F. Cellini, V. Valentini, and D. Verellen, "Multi-object tracking in MRI-guided radiotherapy using the tracking-learning-detection framework," *Radiotherapy Oncol.*, vol. 138, pp. 25–29, Sep. 2019.

[18] C. Dong, H. Li, X. Luo, L. Ding, J. Siebert, and H. Luo, "Proactive struck-by risk detection with movement patterns and randomness," *Autom. Construct.*, vol. 91, pp. 246–255, Jul. 2018.

[19] Z. Jiang, W. Zhou, H. Li, Y. Mo, W. Ni, and Q. Huang, "A new kind of accurate calibration method for robotic kinematic parameters based on the extended Kalman and particle filter algorithm," *IEEE Trans. Ind. Electron.*, vol. 65, no. 4, pp. 3337–3345, Apr. 2018.

[20] L. Ma, "Construction of intelligent building sky-eye system based on multi-camera and speech recognition," *Int. J. Speech Technol.*, vol. 23, no. 1, pp. 23–30, Dec. 2020.

[21] Z. Zhang and F. Cohen, "3D pedestrian tracking and frontal face image capture based on head point detection," *Multimedia Tools Appl.*, vol. 79, nos. 1–2, pp. 737–764, Sep. 2020.

[22] W. Osamy, A. M. Khedr, A. Salim, and D. P. Agrawal, "Sensor network node scheduling for preserving coverage of wireless multimedia networks," *IET Wireless Sensor Syst.*, vol. 9, no. 5, pp. 295–305, Oct. 2019.

[23] J. Yang, B. Jiang, Z. Lv, and K.-K.-R. Choo, "A task scheduling algorithm considering game theory designed for energy management in cloud computing," *Future Gener. Comput. Syst.*, vol. 105, pp. 985–992, Apr. 2020.

[24] M. Quintana, S. Karaoglu, F. Alvarez, J. M. Menendez, and T. Gevers, "Three-D wide faces (3DWF): Facial landmark detection and 3D reconstruction over a new RGB–D multi-camera dataset," *Sensors*, vol. 19, no. 5, pp. 1103–1124. Jan. 2019.

[25] L. Wang, R. Gao, J. Váncza, J. Krüger, X. V. Wang, S. Makris, and G. Chryssolouris, "Symbiotic human-robot collaborative assembly," *CIRP Ann.*, vol. 68, no. 2, pp. 701–726, Jun. 2019.

[26] H. He, G. Liu, X. Zhu, L. He, and G. Tian, "Interacting multiple model-based human pose estimation using a distributed 3D camera network," *IEEE Sensors J.*, vol. 19, no. 22, pp. 10584–10590, Jul. 2019.

[27] F. Zhou, "Dynamic tracking algorithm suitable for intelligent video surveillance," *J. Comput. Methods Sci. Eng.*, vol. 19, no. S1, pp. 157–164, Aug. 2019.

[28] Y. He, X. Wei, X. Hong, W. Shi, and Y. Gong, "Multi-target multi-camera tracking by tracklet-to-target assignment," *IEEE Trans. Image Process.*, vol. 29, pp. 5191–5205, 2020.

[29] W. C. Choi and J. Y. Na, "Development of CCTV cooperation tracking system for real-time crime monitoring," *J. Korea Acad.-Ind. Cooper. Soc.*, vol. 20, no. 12, pp. 546–554, Dec. 2019.

[30] D. J. Taylor, N. D. Smith, P. R. Jones, A. M. Binns, and D. P. Crabb, "Measuring dynamic levels of self-perceived anxiety and concern during simulated mobility tasks in people with non-neovascular age-related macular degeneration," *Brit. J. Ophthalmol.*, vol. 104, no. 4, pp. 529–534, Mar. 2020.

[31] R. Iguernaissi, D. Merad, K. Aziz, and P. Drap, "People tracking in multi-camera systems: A review," *Multimedia Tools Appl.*, vol. 78, no. 8, pp. 10773–10793, Sep. 2019.

[32] Y. Zhang, Y. Zhang, D. Li, K. Liu, and Y. Fan, "Dual-polarized band-notched antenna without extra circuit for 2.4/5 GHz WLAN applications," *IEEE Access*, vol. 7, pp. 84890–84896, 2019.

[33] J.-Y. Chang, S.-M. Hong, D. Son, H. Yoo, and H.-W. Ahn, "Development of real-time video surveillance system using the intelligent behavior recognition technique," *J. Inst. Internet, Broadcast. Commun.*, vol. 19, no. 2, pp. 161–168, Apr. 2019.

[34] T. Tamuhla, L. Joubert, D. Willemse, and M. J. Williams, "SufT is required for growth of mycobacterium smegmatis under iron limiting conditions," *Microbiology*, vol. 166, no. 3, pp. 296–305, Mar. 2020.

[35] V. J. McIver, L. Mattin, G. H. Evans, and A. M. W. Yau, "The effect of brisk walking in the fasted versus fed state on metabolic responses, gastrointestinal function, and appetite in healthy men," *Int. J. Obesity*, vol. 43, no. 9, pp. 1691–1700, Sep. 2019.

[36] R. Sun, J. Qian, R. H. Jose, Z. Gong, R. Miao, W. Xue, and P. Liu, "A flexible and efficient real-time ORB-based full-HD image feature extraction accelerator," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 28, no. 2, pp. 565–575, Feb. 2020.

[37] R. F. Riley, S. J. Walsh, A. J. Kirtane, R. M. Wyman, W. J. Nicholson, L. Azzalini, J. C. Spratt, S. Kalra, C. G. Hanratty, A. Pershad, T. DeMartini, D. Karmpaliotis, W. L. Lombardi, and J. A. Grantham, "Algorithmic solutions to common problems encountered during chronic total occlusion angioplasty: The algorithms within the algorithm," *Catheterization Cardiovascular Intervent.*, vol. 93, no. 2, pp. 286–297, Nov. 2019.

ZHIQING ZHOU was born in Zhejiang, China, in 1978. He received the master's degree in software engineering from Hangzhou Dianzi University, China, in 2007. He is currently working as a Lecturer with the School of Information Engineering, Pingdingshan University, Henan. He has published five articles. His research interests include the Internet of Things and machine vision.

HENG YU was born in Henan, China, in 1984. He received the bachelor's and master's degrees from Henan University, in 2008 and 2011, respectively. He is currently working as a Lecturer with the School of Information Engineering, Pingdingshan University, Henan. He has published nine articles, one of which has been indexed by EI. His research interests include digital image processing and machine learning.

HESHENG SHI was born in Luohe, Henan, China, in 1977. He received the master's degree in computer technology from Xidian University, China, in 2010. He is currently working as an Assistant Professor with the School of Information Engineering, Pingdingshan University, Pingdingshan, Henan. His research interests include software engineering and networking algorithms.