Journal of Computer Technology and Software

ISSN:2998-2383

Vol. 1, No. 1, 2022

Enhancing Efficiency and Accuracy in Video Surveillance Systems: A Target Detection Model Utilizing Open Pose Algorithm

John Doe[1]

Department of Electrical and Computer Engineering, University of California, Irvine, USA

Abstract: Video surveillance plays a critical role in maintaining public security by offering real-time, comprehensive, and easily interpretable image information. While traditional surveillance relies heavily on human operators who are susceptible to visual fatigue, intelligent video surveillance systems have emerged as a solution to enhance monitoring efficiency and effectiveness. This paper explores the development and application of a video surveillance system model leveraging the Open Pose algorithm for target detection. The model's ability to detect and track multiple targets, recognize abnormal behaviors, and assess crowd and traffic flow is highlighted. Experimental results demonstrate significant improvements in monitoring efficiency and accuracy, although challenges remain in preprocessing under unstable environmental conditions. Future work will focus on enhancing the preprocessing module to address these challenges and further improve system performance.

Keywords: Open Pose, Video monitoring system, Virtual electronic fence alarm system.

1. Introduction

Video surveillance is an important carrier of image information. It has the advantages of real-time, containing more information content and intuitive and easy to understand. It is an important part of maintaining social public security. Video surveillance system can not only be used to investigate and collect evidence of illegal and criminal acts, but also an effective means of crime prevention. The use of computer vision related technologies to process surveillance video data is a major development hotspot at present [1]. Potential criminals are often afraid in places covered by surveillance. There is no doubt that video surveillance system in public places is an important guarantee for people's life safety. Many public places need relevant personnel to find and deal with abnormal conditions immediately. At present, most of them rely on surveillance video for real-time monitoring by security personnel. However, people have the physiological phenomenon of visual fatigue and cannot maintain long-term attention. Therefore, only traditional monitoring can play a very limited role in avoiding accidents [2].

Intelligent video surveillance system technology can basically realize the functions of face detection, dangerous area intrusion detection, behavior recognition and so on, and has realized the preliminary application in many fields. W4 real-time visual monitoring system researched by the University of Maryland combines visual tracking and pose analysis. By locating the trunk, limbs and other parts of the human body in the monitoring video, the human body appearance model is established, and the functions of multi-person detection and multi-target tracking are realized outdoors. In addition, the system also considers the factors such as occlusion, which can not only judge whether the human body wears other objects, but also realize the contour segmentation of the human body and objects [3]. In China, the research on intelligent monitoring system started relatively late, but in recent years, driven by the upsurge of domestic artificial intelligence development, domestic research institutions, universities and enterprises have extensive research in the field of intelligent monitoring system, and have achieved rapid development in application [4]. Key research institutions such as Microsoft Research Institute in Asia and the State Key Laboratory of visual and auditory information processing of Peking University have made important research achievements. In particular, the Institute of automation of the Chinese Academy of Sciences has made great progress in relevant research fields [5]. The biological recognition and security technology research center of the Institute of automation has developed an intelligent video monitoring system. The main functions of the system include multi-target detection, tracking and classification of people and vehicles; Recognition of abnormal behavior of moving target; Recognition of abnormal human actions; Face tracking and recognition in surveillance video; Detection of abnormal objects left and lost; Crowd and traffic flow assessment, vehicle counting and congestion alarm, etc.

Human behavior recognition, as the core content of intelligent visual monitoring system, originated in the 1870s [6].. Based on this experiment, researchers found the relationship between movement and bone key points, so they began to study bone key points. Entering the 21st century, human behavior recognition technology is gradually developing at a high speed. However, due to the complexity of human body movements and the influence of the complexity of the environment, the detection method of human behavior is relatively complicated, and its practicability and universality are not high. However, with the development of deep learning, limb detection technology has also been significantly improved, and researchers have begun to use machine learning theory to detect limb key points. The methods of limb detection are divided into two forms: top-down and bottom-up. At the same time, in the research of limb detection, some models of skeleton key point detection have emerged as the times require. For example, the CPM model, which was proposed by Shih-EnWei [7], etc., is a multi-level

network. Its principle is to use the heat map output through the human body image to characterize the position information corresponding to the key points, and grade the output response as the next step. input to the network until the final output is obtained. The Deep Cut method was proposed by Insafutdinov [8] et al. This method uses convolutional neural network to extract joint points in candidate regions to represent key points of the skeleton, and the weight between each key point represents the degree of association, and then optimizes and completes the results. Clustering of skeleton key points. The Alpha Pose model uses a top-down method for detection. The advantage of this model is that in the process of top-down keypoint detection, the accuracy of target detection is greatly improved. This method is developed by Hao-ShuFang [9] and other scholars proposed that the model is mainly composed of three parts: SSTN, NMS and PGPG. Among them, for the problems of incomplete human body, repeated detection and insufficient training samples that are often encountered in traditional SPPE detection, SSTN, NMS, PGPG can be better resolved. The feature extraction method of the Open Pose model is also a convolutional neural network. The model is proposed by scholars such as ZheCao [10]. The model extracts the output of the feature extraction, extracts the confidence map and the affinity domain, and then performs clustering [11]. Analyze the key point information of human skeleton.

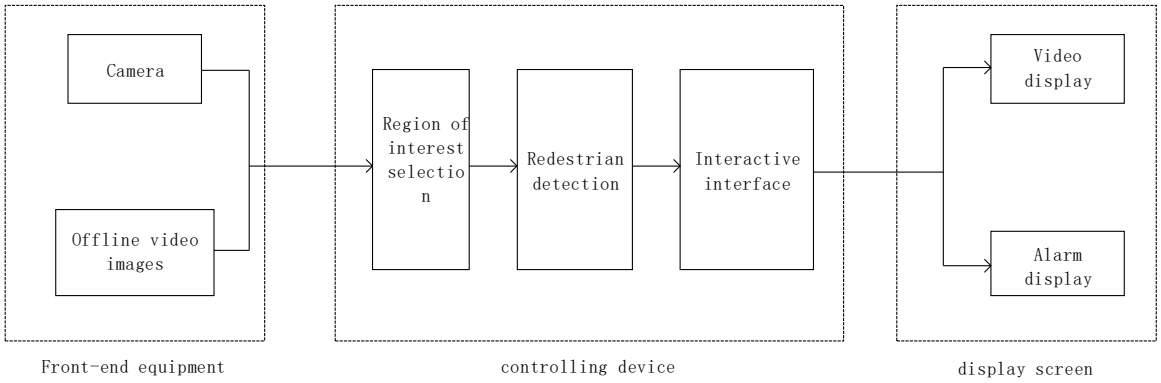

Figure 1. System structure

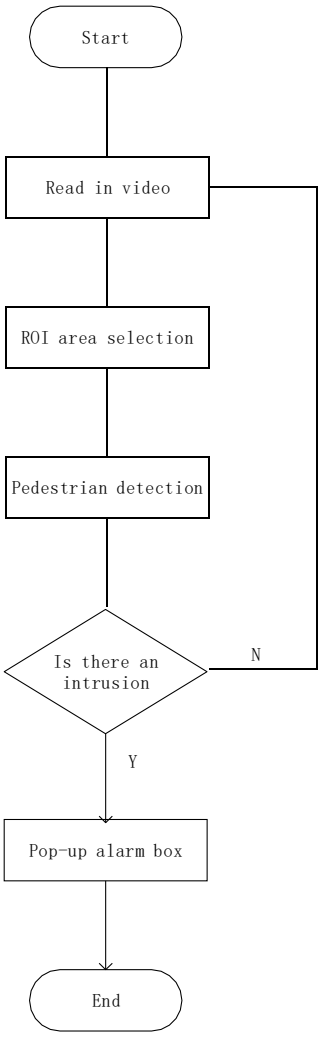

Figure 2. System flow chart

On the premise of integrating the target detection in video surveillance and computational vision, it is applied to the field of electronic fence. This paper proposes a target detection and alarm system model based on video surveillance system. With the support of computer technology, intelligent monitoring and management can be realized. When there are abnormal events in the monitoring field of vision, pedestrians can be identified and warned accordingly to remind the monitoring personnel to deal with them quickly and prevent further accidents. This not only strengthens the monitoring effect, but also greatly improves the work efficiency of the, but also reduces the work intensity of the monitoring personnel to a great extent.

2. System Structure Design

Aiming at the problems of large amount of monitoring and physiological fatigue of staff in video monitoring system, a target detection and alarm system based on video monitoring system is proposed in this paper. The system can set any geometric area independently, recognize the human posture of pedestrians entering the area through the video information collected by the monitoring lens, detect pedestrian intrusion and give an alarm, which plays the role of shock and warning. The system is divided into three modules: front-end equipment, control equipment and display equipment. As shown in Figure 1. The flow chart of the overall system is shown in Figure 2.

2.1. Pretreatment Module

The video image acquisition is completed through the surveillance camera, which is connected with the computer with video acquisition card through video cable. In this design, the network cable is used to replace the professional video cable, but the video clarity collected by the network cable is far less than that of the video cable, so it is necessary to carry out grayscale and other steps to process the video image in high definition. The preprocessing module is shown in Figure 3.

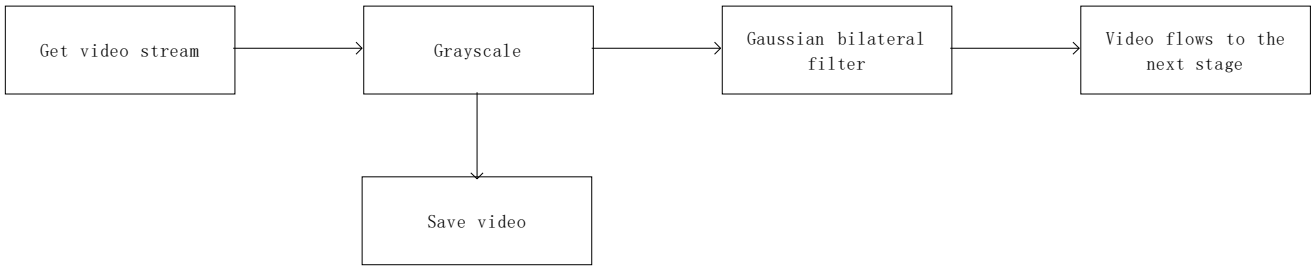

Figure 3. Preprocessing process

The collected original video frame is a three channels color map, the value range of each channel is 0 ~ 255, and there are 256 * 256 * 256 color changes of each pixel, which will cause a lot of storage space. The collected video is stored in the hard disk. Generally, 3G is required for recording one day, and larger capacity is required for multi camera recording. Grayscale processing is to filter out the color information of the image without changing the brightness and other information of the image. This operation can reduce the size of the image without changing the image and facilitate storage.

2.2. Region of Interest Selection

In the field of image processing, region of interest (ROI) is the focus of image analysis, which refers to selecting an image region with arbitrary geometry from the images to be processed. Using ROI to select the target image we want to focus on can simplify the unnecessary processing process and reduce the image processing time.

The video image collected by the video surveillance camera is prone to noise due to the instability of the external environment, which is not conducive to subsequent operation and processing. Therefore, the video image is preprocessed after collection. After eliminating the video noise, selecting the region of interest in the video picture is equivalent to setting the boundary, which is similar to the physical fence. Select the dangerous area, and only detect and recognize the image of the selected area.

When the system starts running, you need to draw arbitrary polygons in the video image as the key area to be monitored. The program flow chart of drawing polygons in key monitoring areas with the mouse is shown in the figure. When you click the left mouse button for the first time, it indicates the starting point of the polygon, and then click the left mouse button in turn as the vertex of the polygon; Each right click represents the end of a polygon drawing. The flow chart of selecting region of interest is shown in Figure 4.

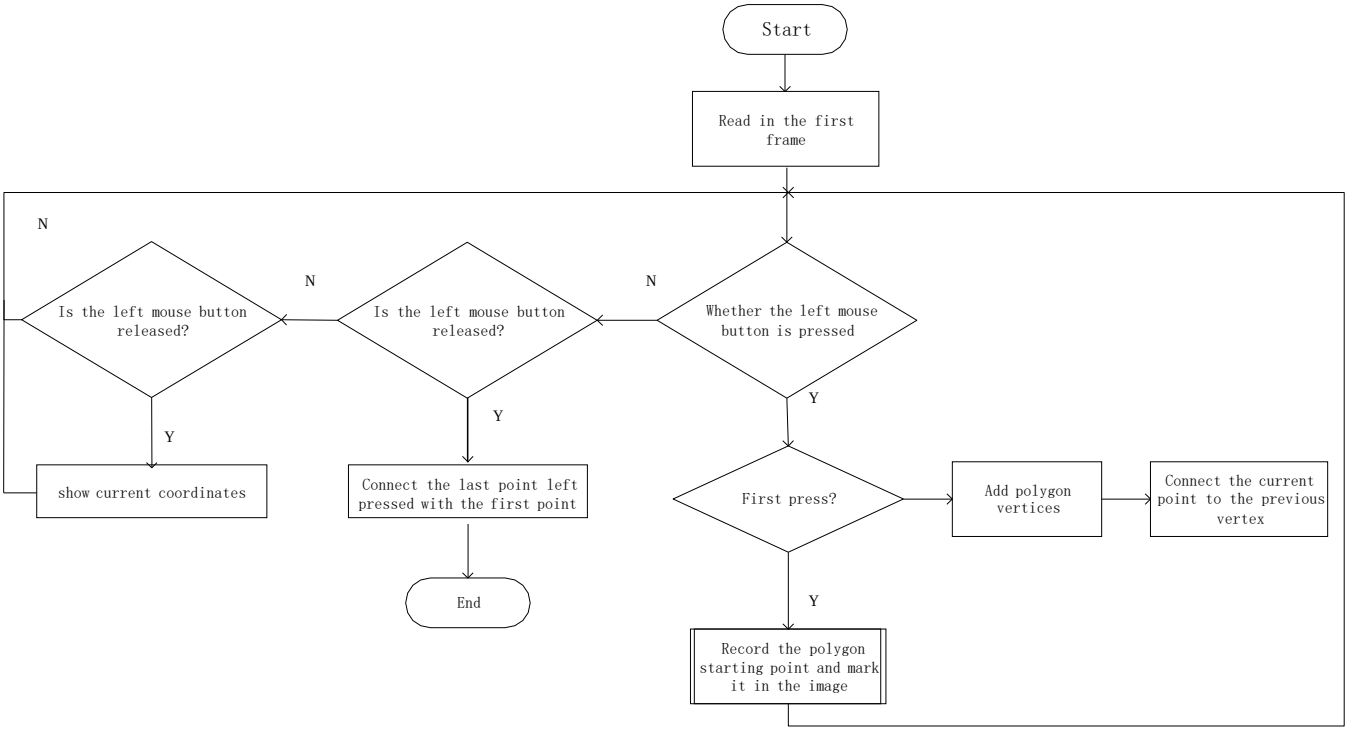

Figure 4. Region of Interest Selection Flowchart

There are two ways to draw an arbitrary polygon with the mouse. One is to fix the first frame or finish drawing in a fixed frame after pausing the video; The second is to complete the drawing without affecting the video playback. As shown in Figure 5.

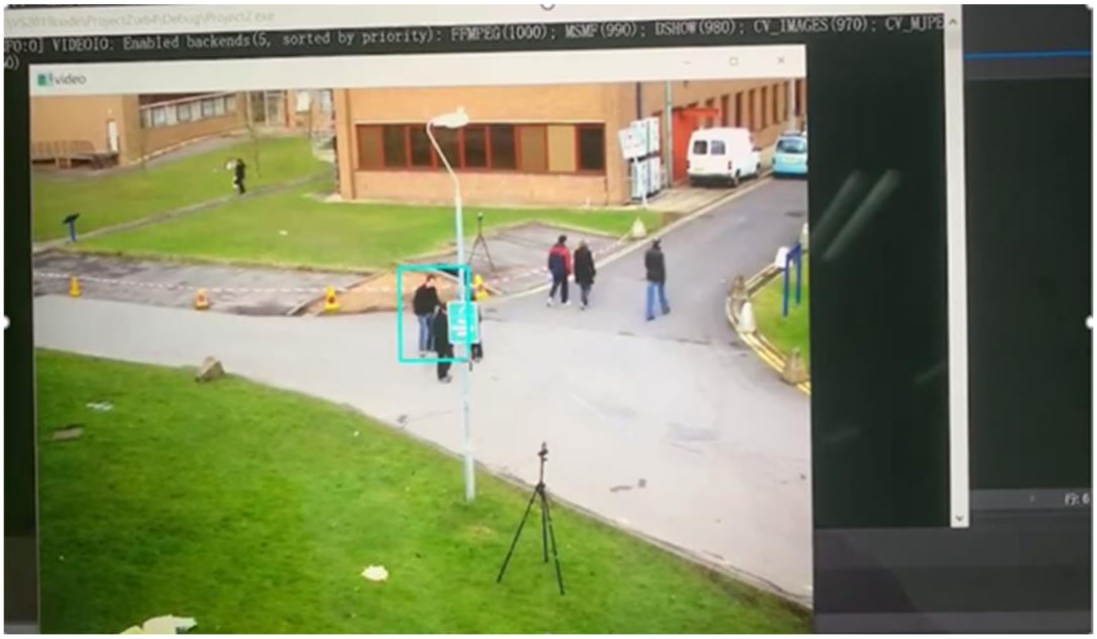

Figure 5. An example of selecting a region of interest

2.3. Pedestrian Detection

Human pose estimation is a significant research topic in computer vision. It plays an important role in many computer vision tasks. For example, action recognition, automatic driving, character tracking, abnormal behavior detection and so on. At the same time, human posture estimation has important application value in the fields of intelligent security, human-computer interaction, virtual reality and so on. In recent years, the related research has made a significant development in human posture estimation. Open Pose is based on convolutional neural network and supervised learning. It can quickly and accurately obtain the skeleton key point information of people in the image. It uses the bottom-up method to detect the skeleton key points of human body. It creatively puts forward the partial affinity field (PAF) representing the correlation degree between joint points, which makes the final clustering efficient and accurate. The test results are shown in the figure 6.

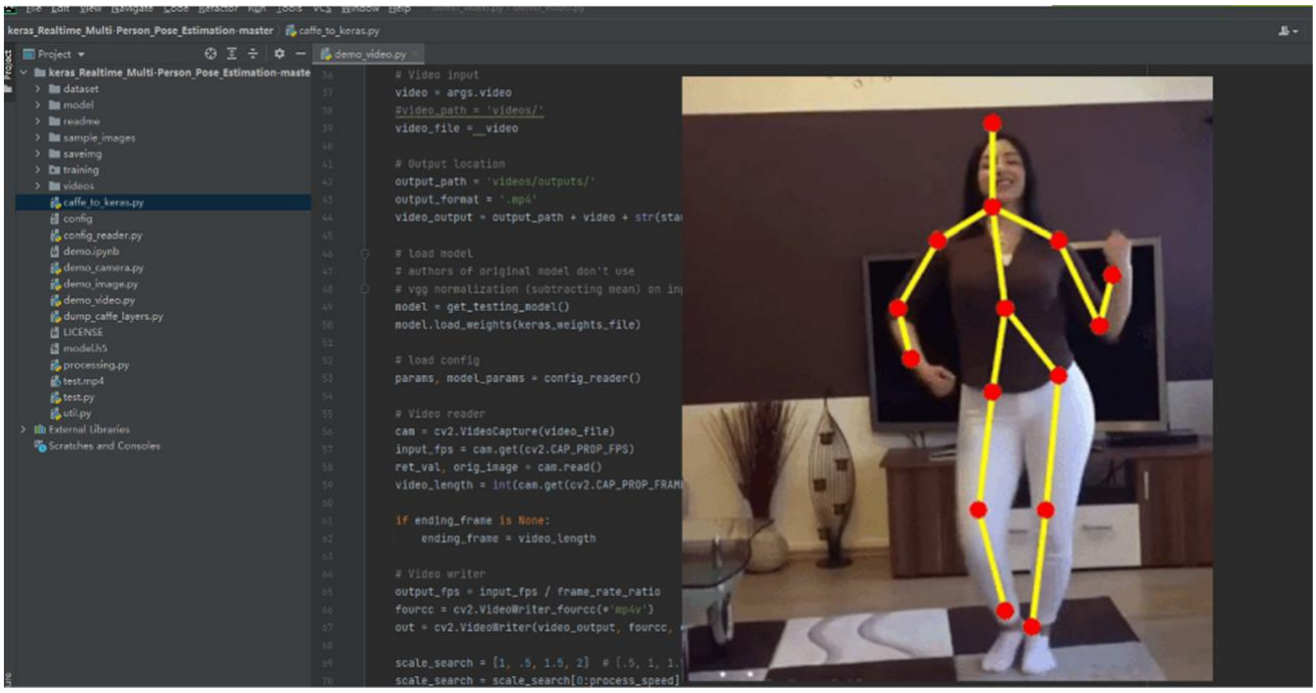

Figure 6. Open Pose detection map

2.4. Visual Display

Based on the functional needs of the system, this paper designs the user interface of the system on QT software. The specific operation functions include drawing ROI area and starting target detection. Each operation interface is shown in the figure. The left position in the middle of the overall picture is the current monitoring picture, and the function key is on the right of the picture. As shown in the figure 7.

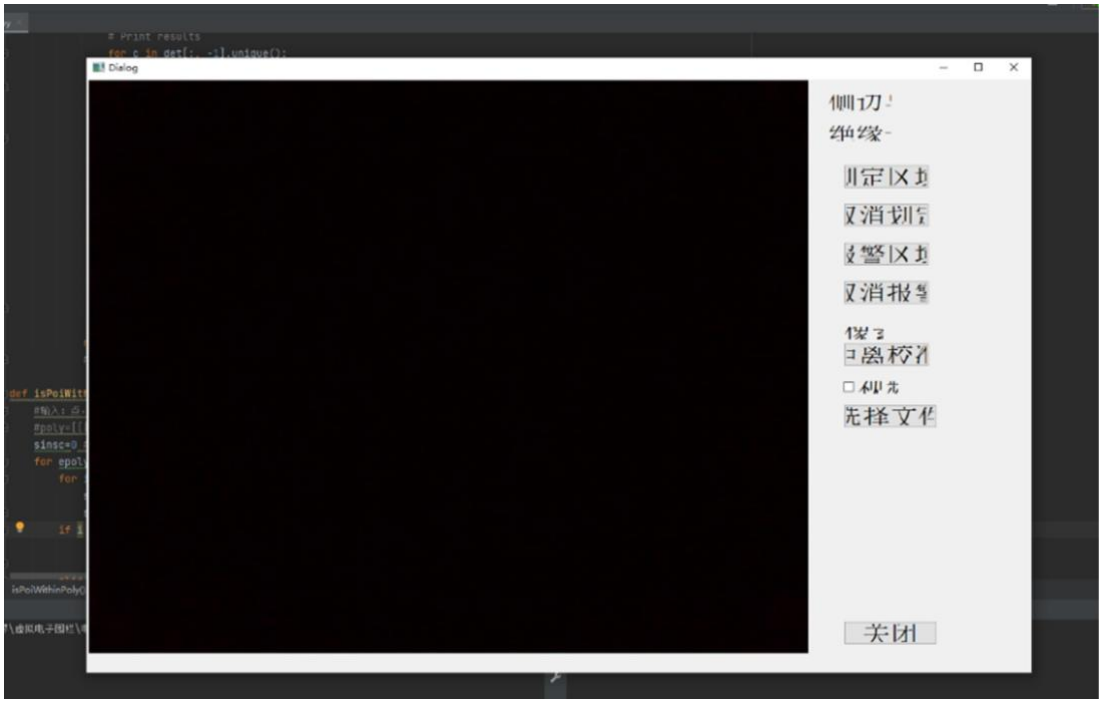

Figure 7. Visual interface display

3. System Development Link

3.1. System Hardware Design

The hardware component of this design is composed of monitoring terminal, small camera, local audible and visual alarm and computer. The small camera is used to collect video images and transmit them to the monitoring terminal. The monitoring terminal is connected with the detection computer through the video line to transmit the collected video stream images to the computer in real time, and the computer processes and detects the video. When pedestrians enter the calibrated image area, the interface will pop up the alarm interface, and the local audible and visual alarm will give an

alarm, which will deter the entering pedestrians. The hardware composition of the system is shown in the figure 8.

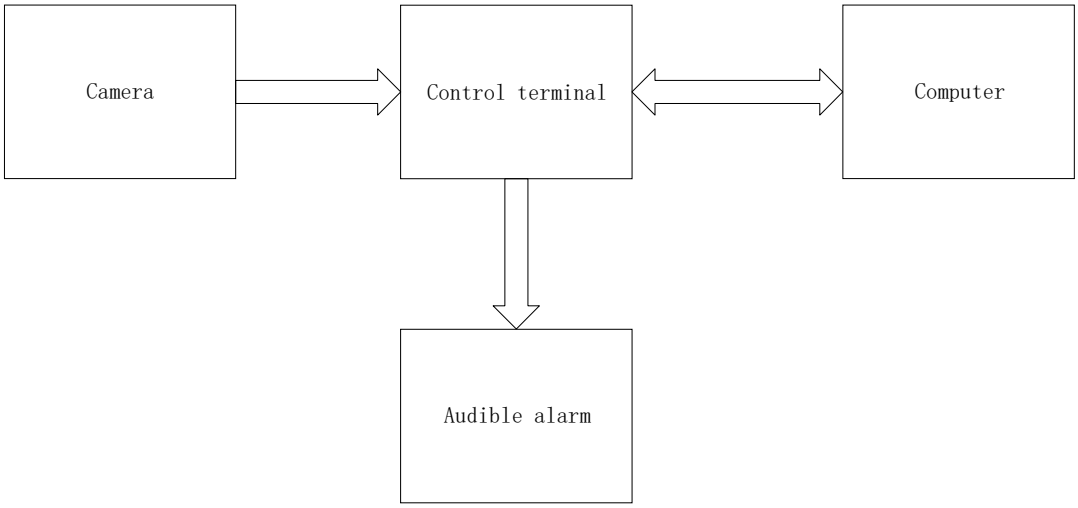

Figure 8. System hardware design diagram

3.2. System Software Development

The software design part of this system is based on windows10 operating system, which completes the configuration of various software environments, including visual studio, OpenCv, CUDA, QT, anaconda, PyCharm and other library files.

4. Analysis of Experimental Results

In view of the large amount of monitoring area in the video monitoring system, the target detection system model based on video monitoring proposed in this paper can effectively improve the efficiency of monitoring video, use Open Pose algorithm to detect the target, and enhance the accuracy of the system model. The practical operation shows that the system model proposed in this paper has good effect., However, the preprocessing module of the system still needs to be improved. How to complete the video image de dithering processing under the unstable factors of the external environment.

References

[1] Xu Lanfei, Qian Xuejun. Three-dimensional simulation research of passenger movements in subway station monitoring system [J]. Railway Computer Applications, 2017，26(6):60-64.

[2] Wu Mengdi. Research on the behavioral characteristics of violent and terrorist crowd monitoring in civil aviation security isolation area [J]. Scientific Chinese,2017 (24).

[3] Yuexin Wu, Zhe Jia, Yue Ming, Juanjuan Sun, Liujuan Cao. Human behavior recognition based on 3D features and hidden markov models[J]. Springer London,2016,10(3): 495–502.

[4] Insafutdinov E, Pishchulin L, Andres B, et al. Deepercut: A deeper, stronger, and faster multi-person pose estimation model[C]//European Conference on Computer Vision. Springer,Cham, 2016: 34-50.

[5] Naji S, Jalab H A, Kareem S A. A survey on skin detection in colored images[J]. Artificial Intelligence Review, 2019, 52(2): 1041-1087.

[6] Pradhan A. Support vector machine-a survey[J]. International Journal of Emerging Technology and Advanced Engineering, 2012, 2(8): 82-85.

[7] Wei S E, Ramakrishna V, Kanade T, et al. Convolutional pose machines[C]//Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2016: 4724-4732.

[8] Insafutdinov E, Pishchulin L, Andres B, et al. Deepercut: A deeper, stronger, and fastermulti-person pose estimation model[C]//European Conference on Computer Vision. Springer,Cham, 2016: 34-50.

[9] Fang H S, Xie S, Tai Y W, et al. Rmpe: Regional multi-person pose estimation[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 2334-2343.

[10] Cao Z, Simon T, Wei S E, et al. Realtime multi-person 2d pose estimation using part affinity fields[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017:7291-7299.

[11] Lv F, Nevatia R, Lee M W. 3D human action recognition using spatio-temporal motion templates[C]//International Workshop on Human-Computer Interaction. Springer, Berlin, Heidelberg, 2005: 120-130.