A Pose Intelligent Monitoring System Based on Message Queuing Telemetry Transport Protocol in the Classroom Background

Jiawei Huang[1, a], Ding Zhou[1, b *]

[1]*School of System Design and Intelligent Manufacturing, Southern University of Science and Technology, Shenzhen 518055, China*
[a]*huangjw@mail.sustech.edu.cn*, [b]*zhoud3@sustech.edu.cn*

Abstract
The prosperity of education contributes to national prosperity and socioeconomic development. The expanding student population in compulsory education makes it technically difficult for teachers to assess students' behavior and learning attitudes. In order to regulate classroom teaching and prevent students from leaving early, missing classes, or cheating on exams, schools have purchased video surveillance equipment and thus made such products enter the education market. However, the current combination of cameras and video recorders has a limited effect due to the lack of intelligent factors. To address this problem, this study is dedicated to introducing an artificial intelligence solution in the education field to deepen education informatization, i.e., using a pose detection algorithm to achieve intelligent monitoring of students' poses in a classroom background through the subscribe and publish mechanism in the Message Queuing Telemetry Transport (MQTT) protocol. The system is able to reflect the students' classroom performance to the teacher, which plays an essential role in promoting the development of teaching programs and improving teaching quality.

Keywords: *Video surveillance, Educational informatization, Pose detection, Message queuing telemetry transport.*

1 INTRODUCTION

Along with the social progress, classroom teaching is in a stage of continuous development and renewal [7] [8], and several primary and secondary schools in China are engaged in building a network of joint teaching to improve the quality of classroom teaching [5]. In response to the current complex teaching condition, many primary and secondary schools in China have adopted campus intelligent management teaching, which constantly promotes the teaching system and the management style and is of great significance in improving the overall teaching quality. Schools use surveillance cameras on a large scale to record the students in the monitoring area. Still, most of these cameras are stored as video files without processing and analysing the video images therein. Most schools currently use 24-hour surveillance cameras. The timely observation of the classroom teaching performance requires constant active monitoring from management personnel, which wastes a lot of resources for storing effective data and labor time in manual detecting, thus justifying the inclusion of the pose intelligence monitoring into the school's monitoring equipment.

In the field of education, automated learning analytics are supposed to record the entire classroom process and provide effective feedback to teachers after pose learning [3]. Jing (2021) argues that the acquisition of classroom behavioral images becomes an urgent need for intelligent teaching and improving teaching quality and therefore proposes a cloud-based database and remote system for real-time image acquisition in English classroom teaching. [1] Kim et al. (2021) analyzed the behavioral and motion changes of high school students in the classroom by exploring the feasibility of applying computer vision technology, and the results demonstrated the feasibility of using gesture detection technology in computer vision in the classroom. However, in the current educational context, most scholars' research only focuses on the pose detection algorithm in the classroom background, ignoring the need for a feasible system built in the real classroom environment. In this paper, against the classroom background, the pose detection algorithm is adopted, and

B. Fox et al. (Eds.): IC-ICAIE 2022, AHCS 9, pp. 858-864, 2023.
https://doi.org/10.2991/978-94-6463-040-4_130

the remote transmission of video image data is realized through the subscription and publish mechanism in Message Queuing Telemetry Transport (MQTT) protocol. The system monitors the pose of students in the classroom by connecting a webcam. [2] [4]

In summary, intelligent monitoring equipment has gradually shared the education market and is dedicated to changing the existing manual analysis of students' classroom performance. This study facilitates the steering group's inspection of teachers' teaching quality by remotely observing students' listening status. Besides, this study further encourages teachers to develop performance improvement plans for students' classroom behavioral habits, and adjust teaching methods according to students' classroom performance in time, thus enhancing teaching quality to a greater extent.

2 SYSTEM ENVIRONMENT AND DATA SOURCES

The hardware environment of the system includes Advanced RISC Machines (ARM) and Artificial Intelligence (AI) host, where the ARM is used to issue subject and data instructions and receive camera information and video image data, and the AI host is used to handle the pose detection against classroom background and data delivery.

(1) ARM

The ARM of this experiment uses the Rockchip RK3288 core-board, mainly for issuing operating instructions. The display module is a 7-inch MIPI LCD screen, supporting backlight brightness infinitely adjustable, equipped with Android OS.

(2) AI Host

The AI host used in this experiment is a personal computer, mainly for receiving and executing instructions. The OS is Ubuntu16.04; the processor is Intel(R) Core (TM) i5-10200H CPU; the computer's main frequency is 2.4 GHZ; the memory 16 G. The Qt Creator software is used as the development platform.

For the data collection in this paper, the data set was acquired by a camera installed in the classroom. And a series of recordings were made for each student, at different time periods, so that each student maintained a different pose at each recording as much as possible, which made it possible for us to process and classify the various data. Ultimately, a total of five days of recording was performed using a camera with an image resolution of 3840×2160. Ten videos were intercepted for testing in the recorded video playback, but some of the data were discarded due to the high similarity of the students' poses. The original images of the video intercepts are shown in Figure 1.

Figure 1: Post detection test image

3 HUMAN POSE DETECTION

Since the system is applied to actual classroom scenarios, a more detailed analysis of students in the classroom background is needed, so this section uses the most common and easily distinguishable actions for detection and analysis, including standing, sitting, and raising hands. These three poses reflect the students' performance in the classroom, thus also reflecting the overall classroom performance, providing teachers with data reference for enhancing their teaching quality. Meanwhile, this section uses OpenPose, a CNN-based human key points detection method, as a framework to analyse and detect the location of human key points and determine the three poses of standing, sitting, and raising hands to realize the classroom pose detection in the video images acquired by the camera. The workflow of classroom pose detection is shown in Figure 2.

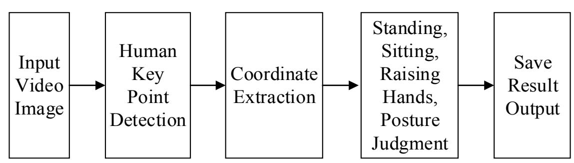

Figure 2: Workflow of classroom pose detection

3.1 *Human Key Point Detection Model*

The COCO keypoint track is currently used as the definitive competition for human key point detection, and the pre-trained model is trained on the COCO dataset. The COCO train set, validation set, and test set contain over 200k images and 250k instances of people labeled with key points. The model is pre-trained by OpenPose, which divides the dataset into a train set and a test set as follows.

- Training. It contains the color image of the human body and the coordinate positions of the key points of the human body. The human body image is obtained from the image compared to a rectangle with a fixed aspect ratio, and the resolution of the rectangle is adjusted by expanding the height and width. Then, the human body image is cropped from the image, and the cropped image is adjusted to a fixed resolution.

- Testing. The model network is tested by first regressing the key points of each human body appearing in the image, predicting the 2D vector domain of the location and orientation of the human limbs based on the partial affinity domain, and repeating the adjustment of the predicted location heat map to locate the final human key points.

3.2 Human Key Point Calibration

In the key point's detection algorithm, the location of the key point changes depending on the pose, and this paper classifies the human poses by identifying the coordinates of the key points corresponding to the poses. According to the above research, this paper calculates the relationship between human key point location and human poses based on Open Pose.

The human key point detection is divided into 18 key points, and their locations are shown in Figure 3. Each key point corresponds to a human body part, as shown in Table 1.

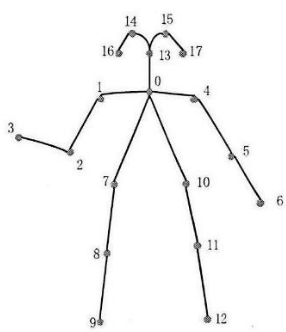

Figure 3: Human key points

Table 1: Definition of human key points

NO	Body Parts	NO	Body Parts
0	Neck	1	Right shoulder
2	Right elbow	3	Right wrist
4	Left shoulder	5	Left elbow
6	Left wrist	7	Right straddle
8	Right knee	9	Right ankle
10	Left straddle	11	Left knee
12	Left ankle	13	Nose
14	Right eye	15	Left eye
16	Right ear	17	Left ear

3.3 Human Pose Detection Results

The Open Pose model is used as a framework for detection, and the three classroom poses of standing, sitting and raising hands are detected by extracting the coordinates of key points and performing operations. The detection effect figure is shown in Figure 4.

Figure 4: Post detection results

In a real classroom background, the standing, sitting, and hand-raising poses detected by the experimental results are identified with boxes, and labels are added. In order to achieve remote monitoring of the system, the MQTT protocol is used for remote transmission of the acquired video image data; therefore, the system is developed for requirements.

4 SYSTEM DEVELOPMENT AND VALIDATION

For this design system, the CPU is expected to be able to realize the video decoding, and bring users clear visual enjoyment as a more efficient HD compression format. From the specific functions of the system, the traditional MCU is not suitable for the system due to the lack of arithmetic power for video coding and decoding. Hence, Rockchip RK3288 main control unit with efficient video coding and decoding is selected to design the system.

4.1 System Scheme

The system uses Rockchip RK3288 as the master control, chooses Hikvision cameras for video shooting, and adopts Network Video Recorder (NVR) for multi-camera management. The design uses the MQTT protocol for data transmission between RK3288 and AI host, sending instructions through RK3288, and the AI host receives the instructions and completes the corresponding instruction requirements; and sends the processed data back to the RK3288 master control board.

The scheme requires installing Hikvision cameras in the classroom and connecting them to the NVR through the network interface, which controls the multiple cameras. Then the MQTT debugging software developed on the AI host connects the IP address of the NVR, and transplants the MQTT debugging software to the RK3288 master control board to establish the data transmission between the RK3288 master control board and the AI host. The RK3288 master control board is responsible for sending instructions and receiving data, while the AI host is in charge of receiving instructions and detecting the pose of classroom video images. After detection, the image data will be sent to the RK3288 master control board through the MQTT protocol, forming a classroom pose detection system. The system chart is shown in Figure 5.

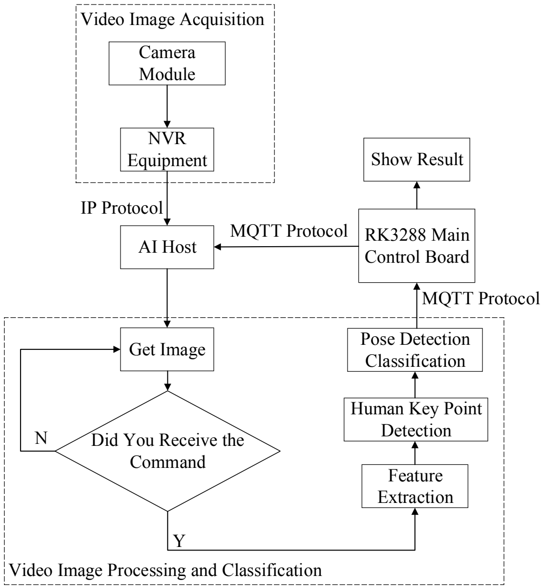

Figure 5: System chart

4.2 *MQTT-based Remote Capture*

The MQTT protocol used in this paper is a relatively easy communication protocol for message transmission, which is also a lightweight transmission with two types of message transmission modes: publisher mode and subscriber mode. The MQTT protocol is extremely suitable for IoT projects and APPs with low network traffic because of these features [6].

The MQTT protocol is used in publishing and receiving data to realize the remote operation of this system. The NVR and MQTT servers are first configured by connecting the AI host to the existing wireless network. Then the MQTT server is connected, and instructions are sent to the upper computer through the ARM. After receiving the instructions, the upper computer will finish capturing or recording the surveillance cameras in the NVR. After executing the capturing or recording, the upper computer will save the data in JSON format and send it to the ARM.

4.2.1 MQTT protocol publishing and receiving

The process of publishing and receiving the MQTT protocol in this system is shown in Figure 6, with the following steps.

- Configuring the NVR and connecting it to the AI host and the ARM RK3288;
- Connecting the MQTT server and connecting both the AI host and the ARM to the MQTT server to establish the connection and data transfer between the two ends.
- Determining the format of communication and data sending from the device by sending topics on the ARM and subscribing to them on the AI host;
- Issuing data instructions by the ARM and receiving data instructions by the AI host;
- The AI host, which receives the data instructions, executes the instructions, and applies Base64 data encoding to the stored pose detection video images; and
- The AI host sends the camera information and image data to the ARM.

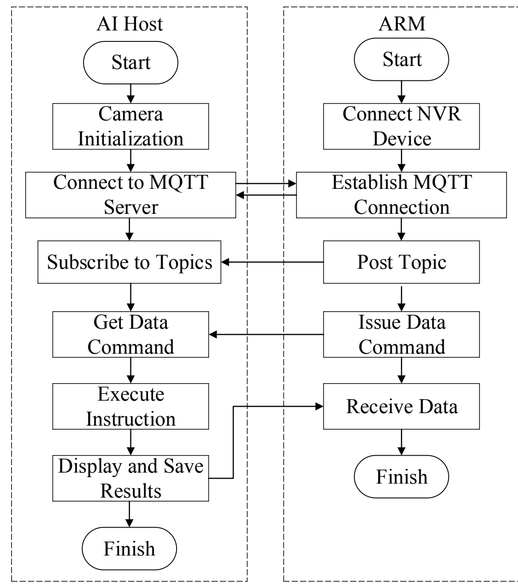

Figure 6: MQTT protocol publishing and receiving

4.2.2 MQTT-based image capture

The upper computer receives the capture instruction from the MQTT server through the established MQTT connection, and the capture process is shown in Figure 7.

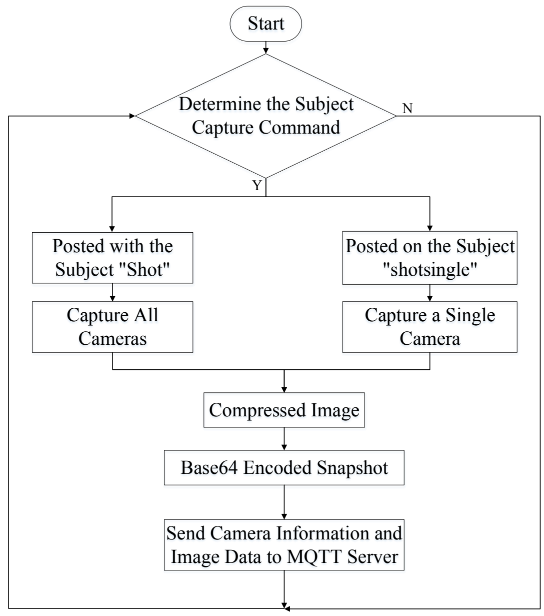

Figure 7: Data released by the upper computer

The instructions sent from the ARM are divided into two categories: the instruction "Capture All Cameras" and "Capture a Single Camera." The processing steps of the captured images are as follows.

- Reading of the local image file address by the upper computer;
- If the captured video image file is large, the image is compressed so that its size does not exceed 200 MB.
- The upper computer converts the image file to data format by Base64 encoding and sends it to the server via MQTT.

4.3 System Validation

The MQTT publishing instructions are divided into the ARM, the AI host, and the MQTT server.

The ARM is equipped with Android OS, which is mainly for publishing topics as well as data instructions and receiving camera information and video image data from the AI host through the MQTT server, as shown in Figure 8.

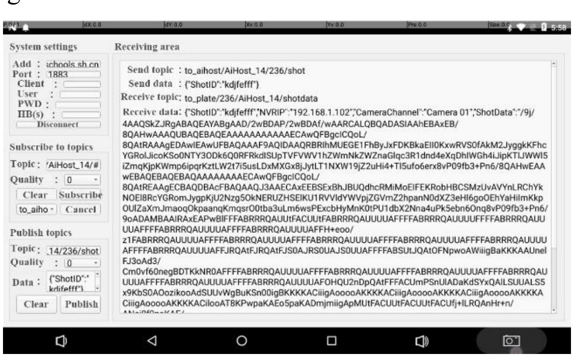

Figure 8: ARM data display

The AI host is mainly used to process instructions issued by ARM, such as camera recording, camera capture, and pose detection of classroom background, and the processed data is sent to the ARM through the MQTT server, as shown in Figure 9.

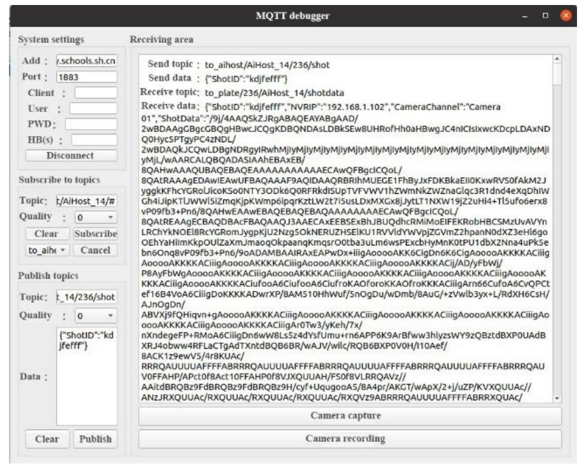

Figure 9: AI host data display

As in Figure 9, the subject and data are included in receiving/sending, expressed as follows.

Subject: to_aihost/{AiHostID}/{HID}/shot

Content: {"ShotID":"", "NVRIP":"", "CameraChannel":"", "ShotData":""}

Each field represents the meaning stated herein.

AiHostID: The unique identifier of the AI host.

HID: The unique identifier of the multi-campus version of the cloud platform.

ShotID: The unique identification of a single capture.

NVRIP: IP address of the NVR.

CameraChannel: The number of camera channels.

ShotData: The data format after the image file is converted by encoding.

The MQTT server serves as a transit station for data, mainly responsible for data transfer between the ARM and the AI host to ensure the accuracy and time-effectiveness of the data.

5 CONCLUSIONS

The system issues video image capture instructions from ARM through MQTT protocol, and the AI host receives the video image capture instructions, saves the video image, and calls the pose detection algorithm to detect and classify the pose in the classroom and identifies the pose detected in the video image. The detected video images are converted into data through the MQTT protocol and published by the AI host to the ARM, which can view the detected video image data and realize the intelligent pose monitoring system in the classroom background. Subscribing to different topics, the configuration of the authority of the upper computer to subscribe to the topics improves the privacy of the upper computer to issue capture instructions and improves the security of the system. The system records students' pose behavior in the classroom by connecting to webcams, objectively evaluates students' pose behavior in the classroom, and assists teachers in enhancing classroom teaching. The system introduces intelligent solutions to the manual monitoring of existing campus cameras, realizes remote intelligent monitoring of students' classroom poses, provides solutions to improve the quality of teaching at the compulsory education level, deepens education informatization, and helps integrate social education with informatization and intelligence.

REFERENCES

[1] Jing, D. A. (2021). Cloud computing database and remote system for real-time image acquisition application in English classroom teaching - sciencedirect. *Microprocessors and Microsystems, 82*.

[2] Kim, H., O'Sullivan, D., Kolykhalova, K., Camurri, A., & Park, Y. (2021). Evaluation of a Computer Vision-Based System to Analyse Behavioral Changes in High School Classrooms. *International Journal of Information and Communication Technology Education (IJICTE), 17*(4), 1-12.

[3] Lam, P. T., Chi, L. P., Tuan, N. H., Dat, N. C., & Son, N. T. (2019). A computer-vision-based application for student behavior monitoring in classroom. *Applied Sciences, 9(*22), 4729.

[4] Liu, X., Zhang, T., Hu, N., Zhang, P., & Zhang, Y. (2020). The method of Internet of Things access and network communication based on MQTT. *Computer Communications, 153*, 169-176.

[5] Qin, L. (2021). Research on the construction of evaluation system of intelligent classroom teaching quality in colleges and universities under the information technology environment. *Journal of Physics Conference Series, 1744*(3), 032164.

[6] Yeh, C. S., Chen, S. L., & Li, I. C. (2021). Implementation of MQTT protocol-based network architecture for smart factory. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 235*(13), 2132-2142.

[7] Zhou, D., Gomez, R., Wright, N., Rittenbruch, M., & Davis, J. (2020). A design-led conceptual framework for developing school integrated STEM programs: the Australian context. *International Journal of Technology and Design Education, 32*(1), 383-411.

[8] Zhou, D., Gomez, R., Davis, J., & Rittenbruch, M. (2022). Engaging solution-based design process for integrated STEM program development: an exploratory study through autoethnographic design practice. *International Journal of Technology and Design Education*, 1-32.