


Research Article

Analysis of Art Classroom Teaching Behavior Based on Intelligent Image Recognition

Chihui Gu ¹ and Yinxing Li²

¹*School of Art and Design, Hunan First Normal University, Changsha 410205, China*

²*School of Humanities and Arts, Hunan International Economics University, Changsha 410205, China*

Correspondence should be addressed to Chihui Gu; gch20220330@163.com

Received 20 May 2022; Revised 27 June 2022; Accepted 30 July 2022; Published 29 August 2022

Academic Editor: Yajuan Tang

Copyright © 2022 Chihui Gu and Yinxing Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To solve the problem of intelligent image recognition in classroom behavior, this paper proposes a fast target detection based on FFmpeg CODEC, extracts MHI-HOG joint features according to the located foreground target area, and finally completes the behavior recognition model through a BP neural network support vector machine joint classifier based on the look-up table. The results are as follows: the motion detection method based on H.264 FFmpeg CODEC video has the highest detection accuracy, which can reach 95%. The foreground detection method takes about 10 ms and saves 90% of the time. The behavior classification and recognition effect of MHI-HOG joint features based on the model has been significantly improved, and the comprehensive recognition rate has reached 95%. The built-in BP neural network support vector machine has 97% accuracy in extracting, classifying, and recognizing the characteristics of a single sample. This study attempts to identify and analyze the class behavior and validate the effectiveness of the collaborative classifiers proposed in this paper to build an intellectual classroom.

1. Introduction

In recent years, with the rapid development of smart hardware and computer technology, the field of computer vision has developed tremendously [1]. Computers are endowed with more functions to complete some monotonous and trivial work or work that human beings cannot complete at present, so as to liberate and improve the productivity. As research deepened, people began to think about computers as human brains, judging, making decisions, and making them smarter [2]. Among the channels of human information acquisition, vision is the most intuitive and also the most comprehensive way to obtain the most information. The vision mentioned here not only refers to the perception of light by human eyes, but also includes a complete set of processes such as visual signal acquisition and transmission, processing and analysis, and memory storage. Therefore, the ability to obtain computer vision is the first step in implementing the machine intelligence. In recent years, computer vision has become a hot field of

research [3]. Today, with the continuous improvement and development of educational recording, broadcasting systems, and artificial intelligence technology, intelligent video recording and broadcasting systems based on video surveillance and recognition technology are developing rapidly, by using the behavior action recognition algorithm based on video sequence images to identify and analyze the behavior action of teachers and students in the classroom scene video. Through students "learning status, it can reflect students" participation in classroom teaching and evaluate the attractiveness of teachers' teaching knowledge to students. Through the analysis of teachers and students' behavior in the classroom, the meeting point between students and teachers in the classroom teaching can be effectively excavated [4]. Through student interactions and points of interest, we can analyze how the teacher's teaching methods and forms of interaction affect students, which can have a positive impact on the curriculum reform, quality improvement, and the development of high-quality talents [5]. Therefore, the research on motion recognition algorithm

based on video has high research value and application prospect not only in multimedia teaching and network resource sharing, but also in video monitoring, civil security, traffic management, and so on [6]. Figure 1 shows the technical scheme of intelligent image recognition system.

2. Literature Review

Due to the relatively late start of computer technology and video surveillance technology, the research on video-based human behavior in China is relatively late. However, with the improvement of infrastructure and the progress of science and technology in recent years, video-based behavior recognition has made rapid development [7]. Scientific researchers can be seen at the top meetings related to CVPR, ICCV, and ECCV [8]. Through continuous dialogue, learning and exchange with foreign high-level researchers, scientific research achievements in this field have become increasingly fruitful. The institutions that conduct research and exploration in this area are mainly some research institutes and universities.

Since the state proposed to vigorously build and develop national high-quality courses in 2003, video resources based on classroom scenes have become more precious [9]. As artificial intelligence and technology have developed in the recent years, the Ministry of education has put forward the concept of keeping pace with the times in 2018. At present, the analysis system for educational scene basically adopts the combination of manual control and infrared induction positioning and tracking to detect and locate human targets [10]. When using this method for video recording, professional personnel need to be equipped to continuously switch scenes and control cameras, which require high labor cost. In addition, the infrared induction method is also easy to receive the interference of external heating objects, resulting in the loss or false detection of target detection and tracking, which has a great impact on the whole system. In this context, the intelligent image analysis system based on video image detection and recognition has attracted the attention of more developers [11]. Video-based detection and recognition systems not only save labor costs, but can also address the classroom interference when manually operated. Video-based classroom recognition and analysis systems generally consist of two parts; teacher behavior analysis module and student behavior analysis module [12]. The video data processed by the teacher behavior analysis module is mainly the scene of the teacher's podium. The teacher is tracked and photographed, and his behavior is analyzed to judge whether the teacher is writing on the blackboard or teaching and questioning PPT courseware. The camera adopts automatic focusing and automatic tracking algorithm for teachers in the shooting process to ensure that the captured video image is clear, smooth, stable, and observable. The teacher tracking and recognition system can complete the automatic recording of classroom video through the intelligent addition, avoid the manual intervention of staff, and ensure that the teachers will not be interrupted in the process of teaching, thus affecting the teachers' teaching status. The Student Behavior Recognition

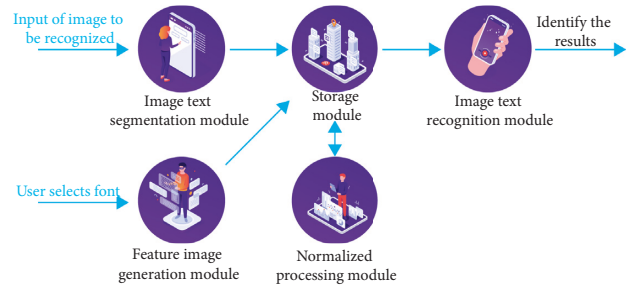


FIGURE 1: Technical scheme of the intelligent image recognition system.

module identifies and analyzes the student behavior in the classroom as a key component. and identifies and analyzes actions in the classroom, such as raising, and answering questions. For students who stand up to answer the questions in class, the camera will take a close-up. When the students finish speaking, they sit down and the camera ends the close-up of the students. When students raise their hands in class, their behavior will be recorded. So as to provide data reference in the follow-up evaluation and analysis of teaching quality. At present, many video surveillance companies have begun to study the intelligent image recognition and analysis system of classroom behavior, and have invested and studied algorithms and products. However, there are great technical difficulties in both the behavior recognition and analysis algorithm and whole machine products. Although the application market of intelligent analysis system which can be applied to classroom scenes and carry out real-time and effective behavior is huge, it still needs further research and development. In recent years, more and more researchers are engaged in the field of human behavior recognition based on Video [13, 14]. In order to recognize the human behavior accurately and efficiently, a large number of algorithms have been proposed. Due to the large number of video scenes and complex behavior background, human behavior recognition is still a challenging research topic. At present, there is no integrated solution to solve human behavior in the whole scene [15]. Therefore, the research focus of this paper is the analysis of human behavior in the classroom scene, trying to identify human behavior efficiently and accurately.

3. Research Methods

3.1. Moving Object Detection Based on the Video

3.1.1. Video Image Preprocessing. Generally, it is not used to directly extract images from video sequence frames for human behavior recognition. Because there will be useless information such as complex background in the picture, and the human target only exists in a part of the image. In addition, the illumination intensity or target shadow will also interfere with the algorithm, which is not conducive to the recognition of human behavior. Therefore, it is necessary to preprocess the input video sequence frames. At present, image preprocessing mainly includes image gray transformation, filtering, enhancement, and other methods. These

methods can improve the original input image in some aspects, remove redundant information, and retain or enhance useful information, and image gray conversion. At present, the mainstream cameras are color cameras, and the video sequences collected are color images. Although color images contain more detailed information and more complete information, they also contain more noise and redundant information. Color image is composed of color information and brightness information. Usually, when processing color images, the algorithm is very complex and the computational complexity increases, it also significantly reduces the real-time performance of the system. At present, there are several commonly used methods to realize RGB image grayscale, such as average grayscale method, weighted average grayscale method, and maximum grayscale method. In the intelligent classroom image recognition and analysis system in this paper, the video stream of the camera is bound to the video processing subsystem (VPSs) module. In this module, video is stored in the form of YUV (4:2:0) [16].

Image denoising: for video images, random noise is often accompanied in the process of acquisition and transmission, which affects the quality of the image. The noise in the image will affect the details of the object, resulting in image distortion. It affects the accuracy of moving object detection and motion recognition. Therefore, noise removal is a necessary link in the image preprocessing. It is currently the most widely used image removal tool: neighborhood mean filter, neighborhood median filter, and Gaussian filter. Image enhancement: the purpose of image enhancement is to enhance the interesting features in the image and suppress the irrelevant features. There are two kinds of commonly used image enhancement methods: methods based on frequency domain and methods based on spatial domain. In the frequency domain, the image is regarded as a two-dimensional signal, and the image can be Fourier transformed. If the image signal is passed through the low-pass filter, the noise in the image can be removed, while the image edge and other information can be enhanced through the high-pass filter to repair the blurred details in the image. Image enhancement in spatial domain is a kind of method that uses more methods. This kind of method is to carry out pixel level operation in spatial domain and complete the processing of the original image through spatial transformation function.

3.1.2. Common Moving Target Detection Algorithms.

Prospect motion target detection is commonly used in video surveillance collar; the aim is to extract the moving target from the complex background from the video sequence captured by the camera. Fast and effective target detection is very important for target tracking, classification, behavior analysis, and other processing, which directly affects the final result. At present, according to whether the background is transformed or not, detection of moving objects is divided into two types: detection of moving objects under a static background and detection of moving objects under a dynamic background. In this subject, we mainly study the situation of static background or small-scale movement of background. In video image research, we call the pixels that

do not move or change as background pixels, and the pixels in the moving area as foreground pixels.

3.1.3. Moving Object Detection Algorithm Based on FFmpeg CODEC.

In real life, due to the limitation of network conditions or storage space, the actual video sequences are basically compressed and encoded video sequences. At present, the most mainstream video coding standards are H264 and MJPEG [17]. H.264 encoded video has gradually become the most widely used video encoding and decoding standard due to its low bit rate, high image quality, error tolerance, and network adaptability. Combining the features of pixel domain and compressed domain can improve the accuracy of target detection. Get a more accurate traffic target [18, 19]. It also uses a multilayer sensor to capture a moving target. A multilayer perceptron (MLP) is a forward artificial neural network that maps input vectors into a set of output vectors. It is a directional graph consisting of multilayer nodes fully connected to the next layer. Each node except the input node is a neuron (or processing unit) with a nonlinear activation function, see Figure 2.

3.2. Human Behavior Feature Extraction. Research into human behavior has become an important area of video processing. The following introduces several human behavior features commonly used in behavior recognition: hog feature, LBP feature, MEI and MHI feature, and optical flow field feature [20].

Hog feature (histogram of oriented gradient) was first proposed to solve the problem of pedestrian detection. The extraction method of hog feature is to divide the image to be processed into uniform grids, and then calculate the local direction gradient histogram in each grid area to obtain the statistical information of the gradient in the image, because the gradient mainly exists at the edge of the object. In video image processing, hog feature is a common feature descriptor. The following is an example of extracting the hog feature of gray image:

- (a) Calculate image gradient information. The image obtained by deriving in the two directions can not only capture the contour, texture, and other information of the object in the image, but also weaken the influence of illumination. Using one-dimensional template and convolution operation with the image, the gradient of the image in the x and y directions can be obtained. The mathematical expression of this process is shown in

$$\begin{aligned} G_x &= H(x+1, y) - H(x-1, y), \\ G_y &= H(x, y+1) - H(x, y-1). \end{aligned} \quad (1)$$

G_x , G_y represent the gradient map of the image in the x and y directions, respectively. $H(x, y)$ represents the pixel value of the image at the pixel point (x, y) . According to the gradient map of the image in two directions, the gradient map of the entire image

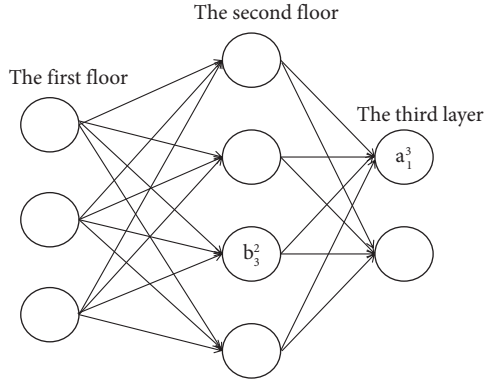


FIGURE 2: Forward propagation of multilayer perceptron.

can be calculated. The calculation method of the gradient size and direction of each pixel is as follows:

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2}, \quad (2)$$

$$\alpha(x, y) = \tan^{-1}\left(\frac{G_y(x, y)}{G_x(x, y)}\right). \quad (3)$$

$G(x, y)$ represents the gradient size of the pixel located at (x, y) in the image, and $\alpha(x, y)$ represents the gradient direction there.

- (b) Divide the image into cells. Divide the image evenly into several cells, each containing 8×8 pixels. The gradient information of pixels in each cell in the image is counted, respectively [21, 22]. See Figure 3, the gradient direction information of pixels in each cell 1 is counted, and a gradient direction histogram is obtained for each cell, which can be represented by an 8-dimensional feature vector.
- (c) *Feature Extension*. Merge the initially divided cells to form a larger cell. By concatenating the eigenvectors of the original cells, the hog feature of the block is obtained [23].
- (d) *Gradient Intensity Normalization*. If the spatial area of image acquisition is large, and there are changes in local illumination in the image, or if the difference between the front and the background changes significantly, the intensity of the resulting gradient will be significantly different. Normalization can compress the edges of the shadow and illumination regions in the image.
- (e) Collect hog characteristics. Taking the detection window size of 120×68 as an example, the cell size is 8×8 pixels, the cell block size is 2×2 cells, the step size is 1 cell, and the number of cell blocks is 14×7 , then the dimension of hog feature is: $(2 * 2) * 9 * (14 * 7) = 3528$.

For HOC features, the extracted edge and gradient features can well express the local shape of the image, quantify the gradient direction by dividing the direction grade, and extract it locally [24, 25]. Therefore, it has good invariance in geometry, and small-scale transformation or

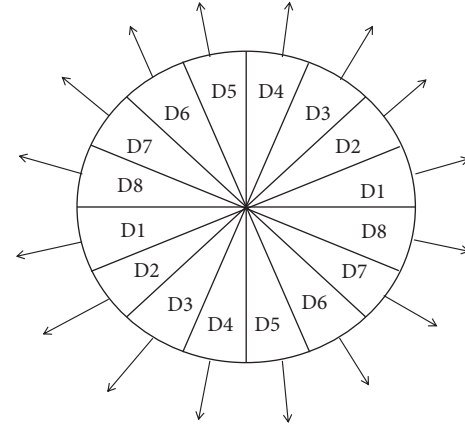


FIGURE 3: Grading diagram of gradient direction.

rotation has little impact on the result. The biggest disadvantage of this algorithm is its slow speed, poor real-time performance, and sensitive to the occlusion of the target.

MEI and MHI characteristics: Davis and bobick first used the contour information of human body to describe the human action. In their method, Motion Energy Image (MEI) and Motion History Image (MHI) are used to display the information about the behavior of the target object. Finally, a behavior classifier was used to recognize and classify the action. Both MEI and MHI feature maps can be called motion templates. The concept of motion template was first proposed by MIT Laboratory in the United States. The acquisition process of MEI and MHI is as follows: according to the motion region detection method introduced in the previous chapter, the motion region in the image can be successfully extracted. In the motion energy map, mark the region where the motion has occurred within a certain time to obtain the motion energy map. For the motion history map, the time of the action of the moving target is expressed in the form of image brightness. The latest action is the brightest. The longer the action is from the current time node, the darker the brightness is. The following is the mathematical representation of the construction process of motion history map, as shown in

$$H_\tau(x, y, t) = \begin{cases} \tau, & \text{if } (\psi(x, y, t) = 1), \\ \max(0, H_\tau(x, y, t-1) - \delta), & \text{otherwise,} \end{cases} \quad (4)$$

$$\psi(x, y, t) = \begin{cases} 1, & \text{if (foreground),} \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

$H(x, y, t)$ represents the brightness value of position (x, y) at time t . t represents the duration and the time range of motion. $\psi(x, y, t)$ represents the update function. If the pixel position of the image is the moving area, $\psi(x, y, t) = 1$, otherwise it is 0. δ represents the decay parameter and represents the brightness loss value of the non-moving area at the current time.

Optical flow field: Gibson first proposed the concept of optical flow in 1950. The motion of an object in a three-dimensional space is called a motion field. When an object

moves in three-dimensional space, the projection in two-dimensional plane is called an optical flow field. The optical flow field represents the speed of pixel motion in a two-dimensional image of an object moving in three-dimensional space. Through the change of the pixel in the video sequence frame and the correlation with the object in the adjacent frame, it is a method to calculate the motion information of the object in the adjacent video sequence image. Optical flow fields can be simply divided into three types: phase-based methods, matching methods, and energy-based methods. The core of optical flow research is to determine the motion information of each pixel position according to the correlation between adjacent frames in video sequence frames and the information of pixel changes. The optical flow method is adopted for feature extraction, and the following three preconditions are adopted by default:

- (1) The sharpness of the two adjacent frames does not change drastically
- (2) The motion of objects in two adjacent images will not change dramatically
- (3) The motion of pixels is consistent in the same subgraph

$|\vec{u}| = \sqrt{u^2 + v^2}$ represents the instantaneous speed of motion, and $\alpha = \tan^{-1}(v/u)$ represents the direction of motion.

According to the three assumptions of the optical flow usage scenario, the following relational expression exists:

$$I(x, y, t) = I(x + dx, y + dy, t + dt). \quad (6)$$

$I(x, y, t)$ represents the gray value of the pixels (x, y) in the current video sequence.

According to the continuity of the object in the image in space and the continuity of the video frame in time, the first-order Taylor expansion of the above formula is obtained.

$$I(x + dx, y + dy, t + dt) = I(x, y, t) + \frac{\partial I}{\partial x} dx + \frac{\partial I}{\partial y} dy + \frac{\partial I}{\partial t} dt. \quad (7)$$

Formula (8) is obtained by sorting.

$$I_x dx + I_y dy + I_t dt = 0. \quad (8)$$

Because of $u = dx/dt$ and $v = dy/dt$, divide both ends of the equation by d to obtain.

$$I_x u + I_y v + I_t = 0. \quad (9)$$

According to the hypothesis, the motion of objects in space is consistent in a very small region. Therefore, formula (10) exists in a small range.

$$\begin{bmatrix} I_{x1} & I_{x1} \\ I_{x2} & I_{x2} \\ \cdots & \cdots \\ I_{xn} & I_{xn} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_{t1} \\ I_{t2} \\ \cdots \\ I_{tn} \end{bmatrix}. \quad (10)$$

We have $A\vec{u} = B$. The mathematical calculation formula of optical flow field is shown in

$$A\vec{u} = B \Rightarrow \vec{u} = (A^T A)^{-1} A^T B, \quad (11)$$

$$A^T A = \begin{bmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_y I_x & \sum I_y^2 \end{bmatrix}.$$

According to the application premise of optical flow field and the above formula, the characteristics of optical flow field are local characteristics. Through two consecutive video sequence frames, the operation information of the object in the video can be calculated, including the speed and direction of motion. In terms of global characteristics, the optical flow field can to some extent reflect the motion information of the object. The main disadvantage of this method is the relatively large number of calculations that are not suitable for real-time systems.

MHI-HOG joint features. According to the introduction of common behavior recognition features, a single feature cannot fully represent the behavior features. For example, although the hog feature can well describe the human behavior, the hog feature is still based on the feature of a single frame image. For example, the image of a frame in the middle process of raising and lowering two actions is basically the same, but the action cannot be distinguished from the hog feature. The information of behavior in the time dimension is not taken into account. The motion history map takes into account the information of action sequence in the time dimension. It can not only display the position of the action, but also reflect the sequence of the movement. For example, it can be easily judged that a frame is in the process of raising or lowering hands through the motion history map. Therefore, in this system, it is considered to combine the motion history map and hog features to form a new feature, and then use the classifier for motion recognition. Firstly, through the foreground moving target detection method, according to the motion vector image obtained in the process of video stream decoding, and then through image filtering and morphological processing, the moving target region is extracted. Finally, the hog feature of the motion history map is extracted as the feature of behavior.

3.3. Human Behavior Recognition and Classification Method.

The design and selection of human behavior recognition classifier is the most important part of the whole system. At present, the mainstream behavior recognition and classification methods include: template matching method, artificial neural network, K-nearest neighbor algorithm, and support vector machine.

Template matching is the most representative algorithm in the field of pattern recognition. Its core idea is to compare the feature vector of the image to be recognized with the feature vector saved in the template, select the one closest to the feature vector to be recognized in the template, and then classify the image to be recognized into this

category. In the template matching method, the standard template library should be built-in advance. The quality of the standard template library directly affects the accuracy of recognition.

Artificial neural network (ANN) has been the focus of academic research since 1980s. Based on the theory of bionics, it digitally simulates the interaction between brain neurons to obtain the artificial neural network model. BP neural network belongs to multilayer feedforward neural network. In order to carry out effective classification and recognition, the network must be trained and learned. Its self-learning is realized through error back propagation in the training stage. K-nearest neighbor algorithm (KNN) is an example-based method, which is the simplest classification method in classification and recognition technology. In the process of classification and recognition, if the classification categories of all training samples are statistically recorded in the training stage, the classification of samples can be completed when the attributes of test samples are exactly the same as those of a training sample. In the actual operation process, the selection of K generally does not exceed 20. In the process of KNN classification and recognition, the selected neighbors are samples that have been correctly classified.

Support Vector Machine (SVM) is an algorithm based on the statistical theory. Auxiliary vector machines were first used to solve binary problems. This method has the advantage of small sample size, nonlinearity, and high-dimensional recognition. It can ensure that the model with low complexity still has good learning ability, that is, the model has good adaptability to samples. Support vector machine can ensure that in the case of small training samples, high prediction accuracy in the training set and low prediction error rate in the test set, so as to realize the best compromise between the complexity of the network system and learning ability. Data classification is done by dividing the data into different subspaces. The mathematical expression for the optimal classification of a hyper plane is

$$F(x) = \text{sgn} \left\{ \sum_{i=1}^n w_i^* y_i K(x_i, x) + b^* \right\}. \quad (12)$$

w_i^* and b^* represent an optimal solution to the problem of solving quadratic programming with constraints, and $K(x_i, x)$ represents a kernel function that maps raw function sample data from low-dimensional function space to high-dimensional function space, which can map the data to the linearly separable high-dimensional feature space. In practical applications, the three most commonly used kernel functions are polynomial kernel function, radial basis function kernel function, and sigmoid inner product kernel function.

The mathematical expression of polynomial kernel function is shown in

$$K(x, x_i) = [x^* x_i + 1]^\alpha. \quad (13)$$

The mathematical expression of radial basis kernel function (BF) is shown in

$$K(x, x_i) = \exp\left(-\frac{|x - x_i|^2}{\delta^2}\right). \quad (14)$$

The mathematical expression of sigmoid inner product kernel function is shown in

$$K(x, x_i) = \tanh[v(x * x_i) + c]. \quad (15)$$

The main idea of SVM can be summarized as follows: first, SVM is designed for linearly separable properties. If the sampling properties are linearly inseparable in a small space, the kernel function is used to draw the properties determined from a small dimensional space in a high-dimensional space that can be separated linearly. Second, the implementation of the theory of structural risk reduction leads to the creation of a class hyper plane in a specific space and the optimization of the world with the help of constraints.

4. Result Analysis

4.1. Comparison of Foreground Detection Performance. The effects of these six methods are verified by experiments. As can be seen from the results in Figure 4, the foreground detection accuracy of background difference method, code table method, and vibe method is high. The accuracy of foreground detection by interframe difference method and Gaussian mixture model method is slightly lower. The motion detection method based on H.264 CODEC video used in this paper has the highest detection accuracy, which can reach 95%.

As can be seen from Table 1, the time efficiency of background difference method and interframe difference method is the highest, followed by the foreground detection method based on H.264 encoded and decoded video adopted in this paper. Since the motion vector table is an accessory of video stream decoding and can be obtained directly in the decoding process, the time listed here represents the time when FFmpeg CODEC decodes a frame of video and restores the motion vector table to a motion vector graph. Vibe method, Gaussian mixture model, and code table method take relatively long time.

4.2. MHI-HOG Joint Characteristic Performance. In order to verify the performance of MHI-HOG joint feature, this paper will use SVM classifier for behavior recognition based on the joint feature and hog feature. MHI-HOG joint feature and hog feature have the same dimension, which ensures that the subsequent constructed SVM classifier has the same scale. Figures 5 and 6 show the action recognition results of SVM classifier classification and recognition using MHI-HOG feature and hog feature, respectively.

According to Figure 6 and Table 2, the effect of behavior classification and recognition based on MHI-HOG joint features has been significantly improved, and the comprehensive recognition rate has reached 95%. Therefore, experiments show that the MHI-HOG joint features used in this paper can effectively represent the behavior

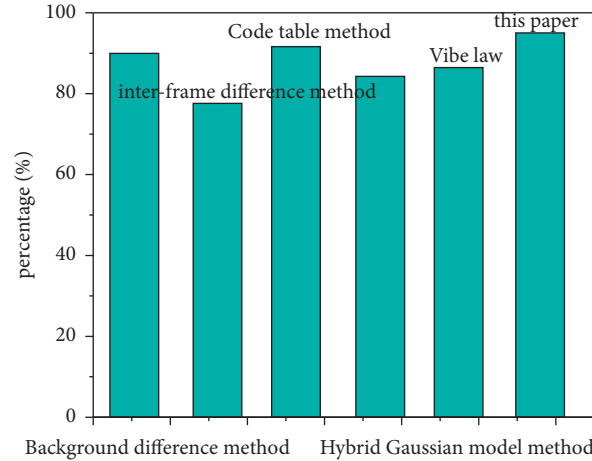


FIGURE 4: Foreground detection accuracy.

TABLE 1: Comparison of time efficiency of foreground target detection.

Method	Background difference method	Interframe difference method	Code table method	Gaussian mixture model	Vibe method	This paper
Single case time (ms)	4.62	4.62	265.2	200.67	105.15	10.36

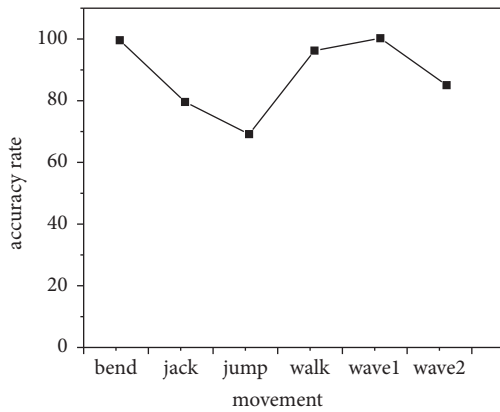


FIGURE 5: Recognition and classification results of SVM classifier based on hog feature.

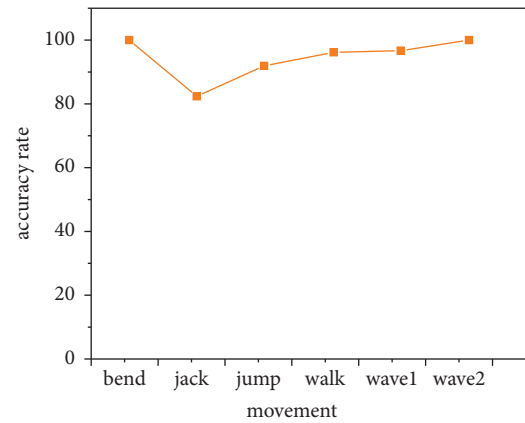


FIGURE 6: Recognition and classification results of SVM classifier based on MHI-HOG joint features.

information. Compared with the traditional hog feature, the processed image itself contains useful information such as the timing relationship of motion actions.

4.3. Performance of the BP Neural Network Support Vector Machine Joint Classifier with MHI-HOG Joint Features. To test the effectiveness of the joint classifier proposed in this paper, the MHI-HOG joint function was used as a behavior recognition function. The BP neural network, support vector machine, and BP neural network support vector machine were used to classify and identify joint classifiers and check the effectiveness of the proposed joint classifier. The results are shown in Figures 7–9 and Table 3.

TABLE 2: Comprehensive recognition rate of two features using SVM classification and recognition.

Features	HOG	MHI-HOG
Correct example	319	338
Total sample		356
Recognition rate	0.90	0.95
Single case time (ms)	55.24	57.87

The BP neural network support vector machine proposed in this paper is slightly increased in time, with a decomposition and classification time of 77.71 ms per sample. But it finally achieved 97% recognition and classification effect, which was the best among the three classifiers.

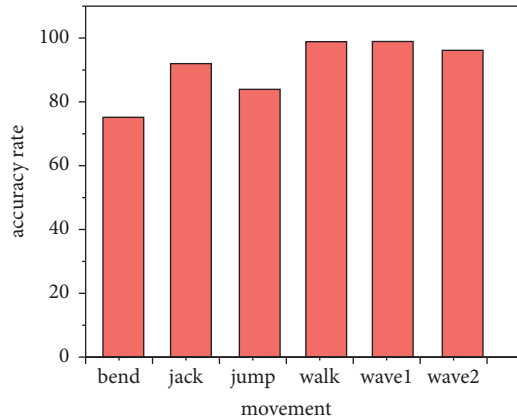


FIGURE 7: Recognition and classification results of BP neural network based on MHI-HOG joint features.

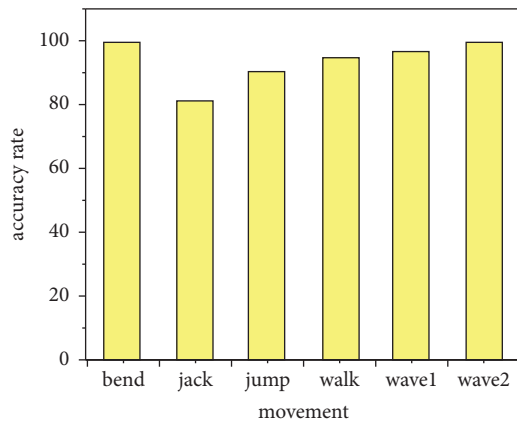


FIGURE 8: Support vector machine recognition and classification results based on MHI-HOG joint features.

Through the analysis of the classification effect of BP neural network and SVM classifier, it can be found that the recognition effect of BP neural network for bend action is poor, and the recognition effect of SVM classifier for Jack action is poor, which affects the final comprehensive recognition rate. With the help of the joint classification and identification of the BP neural network and the SVM, these types of actions can be effectively classified and identified. The experimental data confirm the effectiveness of the joint classifier proposed in this paper.

4.4. Practical Application of the Model. The hardware and software environment required for the implementation of the algorithm and interface used in this paper is shown in Table 4 and Table 5.

The system designed in this paper can complete the collection, processing, analysis, and storage of audio and video by configuring some peripheral chips. The video transmitted through SDI is converted into parallel data in the form of BT1120 through chip G7704. The video transmitted through VGA or HDMI is converted into parallel port data in the form of bt1120 through the chip

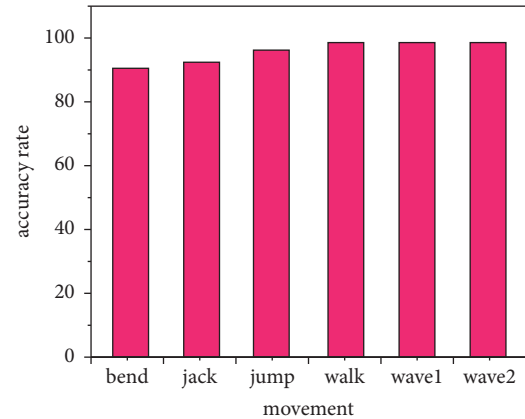


FIGURE 9: Recognition and classification results of BP neural network support vector machine joint classifier based on MHI-HOG joint features.

TABLE 3: Comprehensive recognition rate of the classifier.

Classifier	ANN	SVM	ANN-SVM
Correct example	330	338	346
Total sample	356		
Recognition rate	0.93	0.95	0.97
Single case time (ms)	75.37	57.87	77.71

TABLE 4: Hardware configuration.

Hardware	Dosing
Processor	Intel (R) core (TM) i7-6800K CPU@3.40 GHz
Graphics card	NVIDIA GeForce GTX 1080
Memory	16 GB

adv7842. The network data is transmitted through the network port, and the SATA port completes the storage of video and data. HDMI and VGA interfaces complete the display of human-computer interaction interface.

In this paper, the intelligent image recognition and analysis system of classroom behavior is adopted to identify the main actions in the classroom, see Figure 10. In practice, a behavioral analysis system can be divided into three parts: human-computer interaction, software systems, and data-driven. The human-computer interaction section mainly performs user instruction input and displays the current state of the system, and can pass user command requirements to the software system. The software system is at the heart of all action recognition algorithms. According to the relevant key factors mentioned in this article, it can be divided into four stages: obtaining video transmission, detecting moving targets, unpacking operational features, and taking action to identify the classifier. The information-based section is responsible for recording, accessing, and managing the classroom behavioral information and video information.

For a class video with a length of 40 minutes, the time consumed by the above corresponding operations is shown in Table 6, and the accuracy is shown in Figure 11.

TABLE 5: Software configuration.

Software	Dosing
System	Ubuntu 16.04
Programming platform	PyCharm
Programming language	Python/C
Deep learning framework	TensorFlow
Interface tools	Tkinter
Dependency library	OpenCV, cuDNN, CUDA, Matplotlib

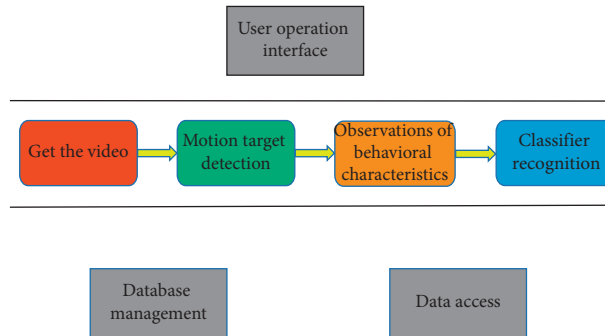


FIGURE 10: Framework flow chart of intelligent image recognition and analysis system for classroom behavior.

TABLE 6: Time consumed by different operations.

Operation	Time consuming	Result
Get video frame	6 minutes and 31 seconds	7200 pictures
Student goal extraction	8 minutes 36 seconds	31 folders, 7200 pictures per folder
Head down and head up recognition in class	1 minute 28 seconds	800 identification pictures and identification result files
Student behavior state recognition	13 minutes and 14 seconds	Identification results corresponding to 31 folders

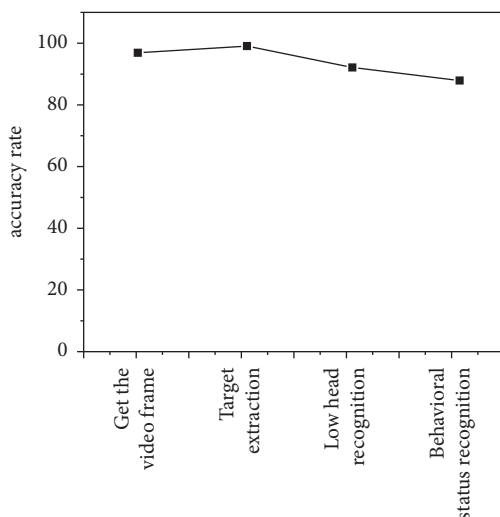


FIGURE 11: Accuracy of different operations.

According to the above results, in practical application, the system designed in this paper has the highest accuracy of 99% in student target extraction, which is much higher than the other similar systems.

5. Conclusion

The video-based classroom behavior analysis system can assist teaching and control the detailed information of the classroom. The video data is also convenient for students to review and reflect after class, which greatly improves the quality of teaching. Under this background, this subject mainly studies the intelligent image analysis system of classroom behavior. Guided by the early process of behavior recognition, this paper makes a theoretical research on three parts: video image motion foreground detection, behavior feature extraction, and behavior recognition classifier. It offers a complete solution for recognizing intelligent images and analyzing the classroom behavior. The main idea of this scheme is fast target detection based on FFmpeg CODEC, extracting MHI-HOG joint features according to the located foreground target area, and finally completing the behavior recognition through a BP neural network support vector machine joint classifier based on the look-up table. The results obtained are as follows:

- (1) The basic moving foreground target detection methods are studied, and the moving foreground target detection method used in this system is proposed. Firstly, several commonly used motion

foreground extraction algorithms are studied and explained theoretically, and the applicable scenarios, advantages, and disadvantages of each algorithm are introduced. Then, experiments are designed to verify the time efficiency and accuracy of several algorithms.

- (2) MHI-HOG feature based on foreground target detection method is proposed. Through the fast motion foreground detection algorithm based on FFMpeg CODEC, the foreground target detection and the update of motion history map are completed, and the hog features of motion history map are extracted. Through experimental verification and analysis on public video data sets, when SVM classifier is used for action recognition, the accuracy of action recognition based on hog feature is 90%, and the accuracy based on the joint behavior feature is 95%.
- (3) This paper constructs a joint classifier based on BP neural network and support vector machine, and uses support vector machine to re-discriminate the discrimination results of the BP neural network. Based on the way of look-up table, on the one hand, it can reduce the scale of SVM classifier; on the other hand, it can find the most suitable classifier for discrimination and improve the accuracy of classification and recognition. On the public video data set, when MHI-HOG feature is used as the recognition feature, the action recognition accuracy based on BP neural network is 93%, the action recognition accuracy based on support vector machine is 95%, and the action recognition accuracy of the joint classifier proposed in this paper reaches 97%, which verifies the performance of the joint classifier.
- (4) Introduced software and hardware platform for intelligent image recognition and analysis of classroom behavior. The experimental analysis is then performed based on the simulated classroom image, and the performance is confirmed on the classroom teacher's actual image. The test results were as expected.

Data Availability

The labeled data set used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The research was supported by the Philosophy and Social Science Project of Hunan Province (Research on Artistic Characteristics and Inheritance and Protection of "Butterfly Mother" Patterns of Miao Nationality in Western Hunan) (No. 18YBA103), the project of Hunan Social Science Achievements Review Committee (Research on Cultural Connotation and Inheritance of Miao Nationality's "Butterfly Mother" Pattern from the Perspective of Regional

Culture Culture) (No. XSP20YBC177), and the "13th Five-Year Plan" of Hunan Education Science in 2018 (Research on the Cultivation and Practice of Innovation and Entrepreneurship Ability of Product Design Students under School-Enterprise Cooperation Mode) (No. XJK18BTW007).

References

- [1] S. Zou, H. Chen, H. Zhou, and J. Chen, "An intelligent image feature recognition algorithm with hierarchical attribute constraints based on weak supervision and label correlation," *IEEE Access*, vol. 99, p. 1, 2020.
- [2] Y. Wu and H. Zhang, "Image style recognition and intelligent design of oiled paper bamboo umbrella based on deep learning," *Computer-Aided Design and Applications*, vol. 19, no. 1, pp. 76–90, 2021.
- [3] H. Zhao, J. Chen, and Y. Lin, "Intelligent recognition of hospital image based on deep learning: the relationship between adaptive behavior and family function in children with adhd," *Journal of Healthcare Engineering*, vol. 2021, no. 3, pp. 1–11, 2021.
- [4] W. Zhao, Y. Guo, S. Yang, M. Chen, and H. Chen, "Fast intelligent cell phenotyping for high-throughput optofluidic time-stretch microscopy based on the xgboost algorithm," *Journal of Biomedical Optics*, vol. 25, no. 6, pp. 1–8, 2020.
- [5] L. Zhang, P. Wang, H. Li, Z. Li, and Y. Zhang, "A robust attentional framework for license plate recognition in the wild," *IEEE Transactions on Intelligent Transportation Systems*, vol. 99, pp. 1–10, 2020.
- [6] M. Zhang, Y. Zhang, Z. Jiang, X. Lv, and C. Guo, "Low-illumination image enhancement in the space environment based on the dc-wgan algorithm," *Sensors*, vol. 21, no. 1, p. 286, 2021.
- [7] T. Lin and X. Liu, "An intelligent recognition system for insulator string defects based on dimension correction and optimized faster r-cnn," *Electrical Engineering*, vol. 103, no. 1, pp. 541–549, 2021.
- [8] Y. F. Zhao, K. Xie, Z. Z. Zou, and J. B. He, "Intelligent recognition of fatigue and sleepiness based on inceptionv3-lstm via multi-feature fusion," *IEEE Access*, vol. 99, p. 1, 2020.
- [9] Z. He, F. Nan, X. Li, S. J. Lee, and Y. Yang, "Traffic sign recognition by combining global and local features based on semi-supervised classification," *IET Intelligent Transport Systems*, vol. 14, no. 5, pp. 323–330, 2020.
- [10] L. Dong, D. Sun, G. Han, X. Li, Q. Hu, and L. Shu, "Velocity-free localization of autonomous driverless vehicles in underground intelligent mines," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 9, pp. 9292–9303, 2020.
- [11] Y. Su, F. Tao, J. Jin, and C. Zhang, "Automated overheated region object detection of photovoltaic module with thermography image," *IEEE Journal of Photovoltaics*, vol. 11, no. 2, pp. 535–544, 2021.
- [12] J. Wang, J. Yu, and Z. He, "Deca: a novel multi-scale efficient channel attention module for object detection in real-life fire images," *Applied Intelligence*, vol. 52, no. 2, pp. 1362–1375, 2021.
- [13] J. Li, J. Su, C. Xia, M. Ma, and Y. Tian, "Salient object detection with purificatory mechanism and structural similarity loss," *IEEE Transactions on Image Processing*, vol. 30, no. 99, pp. 6855–6868, 2021.
- [14] H. Wang, S. Liao, and L. Shao, "Afan: augmented feature alignment network for cross-domain object detection," *IEEE*

- Transactions on Image Processing*, vol. 30, no. 99, pp. 4046–4056, 2021.
- [15] Y. Li, H. Dong, H. Li, X. Zhang, B. Zhang, and Z. Xiao, “Multi-block ssd based on small object detection for uav railway scene surveillance,” *Chinese Journal of Aeronautics*, vol. 33, no. 6, pp. 1747–1755, 2020.
 - [16] B. Gu, R. Ge, Y. Chen, L. Luo, and G. Coatrieux, “Automatic and robust object detection in x-ray baggage inspection using deep convolutional neural networks,” *IEEE Transactions on Industrial Electronics*, vol. 68, no. 10, pp. 10248–10257, 2021.
 - [17] Z. Zheng, L. Lei, H. Sun, and G. Kuang, “Fagnet: multi-scale object detection method in remote sensing images by combining mafpn and gvr,” *Journal of Computer-Aided Design & Computer Graphics*, vol. 33, no. 6, pp. 883–894, 2021.
 - [18] L. H. Wen and K. H. Jo, “Fast and accurate 3d object detection for lidar-camera-based autonomous vehicles using one shared voxel-based backbone,” *IEEE Access*, vol. 99, p. 1, 2021.
 - [19] B. Zhao, Y. Wu, X. Guan, L. Gao, and B. Zhang, “An improved aggregated-mosaic method for the sparse object detection of remote sensing imagery,” *Remote Sensing*, vol. 13, no. 13, pp. 2602–2607, 2021.
 - [20] Z. Luo, X. Lu, and X. Xi, “Eeg feature extraction based on a bilevel network: minimum spanning tree and regional network,” *Electronics*, vol. 9, no. 2, pp. 203–208, 2020.
 - [21] L. Zhong, X. Guo, Z. Xu, and M. Ding, “Soil properties: their prediction and feature extraction from the lucas spectral library using deep convolutional neural networks,” *Geoderma*, vol. 402, no. 5, pp. 115366–115369, 2021.
 - [22] Q. Zhang, Y. Guo, and Z. Y. Song, “Dynamic curve fitting and bp neural network with feature extraction for mobile specific emitter identification,” *IEEE Access*, vol. 9, no. 99, pp. 33897–33910, 2021.
 - [23] B. Zhao, S. Li, Y. Gao, C. Li, and W. Li, “A framework of combining short-term spatial/frequency feature extraction and long-term indrnn for activity recognition,” *Sensors*, vol. 20, no. 23, pp. 6984–6988, 2020.
 - [24] J. Gao, Y. Hong, B. Hong, X. Li, A. Jia, and Y. Qu, “A method of feature extraction of position detection and weld gap for gmaw seam tracking system of fillet weld with variable gaps,” *IEEE Sensors Journal*, vol. 21, no. 20, pp. 23537–23550, 2021.
 - [25] L. Liang, X. Ding, F. Liu, Y. Chen, and H. Wen, “Feature extraction using sparse kernel non-negative matrix factorization for rolling element bearing diagnosis,” *Sensors*, vol. 21, no. 11, pp. 3680–3685, 2021.