Intelligent Traffic Video Retrieval Model based on Image Processing and Feature Extraction Algorithm

Xiaoming Zhao[1], Xinxin Wang[2]*
School of Electrical Engineering and Automation, Luoyang Institute of Science and Technology, Luoyang 471000, China[1]
School of Computer and Information Engineering, Luoyang Institute of Science and Technology, Luoyang 471000, China[2]

***Abstract*—Intelligent transportation is a system that combines data-driven information with traffic management to achieve intelligent monitoring and retrieval functions. In order to further improve the retrieval accuracy of the system model, a new retrieval model was designed. The functional requirements of the system were summarized, and the three stages of data preprocessing, feature matching, and feature extraction were analyzed in detail. The study adopted preprocessing measures such as equalization and normalization to minimize the negative effects of noise and brightness. Based on the performance of various algorithms, the distance method was selected as the feature matching method, which has a wider applicability and is better at processing bulk data. Next, the study utilizes Euclidean distance method to extract keyframes and divides the feature extraction into three parts: color, shape, and texture. The methods of color moment, canny operator, and grayscale co-occurrence matrix are used to extract them, and ultimately achieve relevant image retrieval. The research conducted multiple experiments on the retrieval performance of the model, and analyzed the results of retrieving single and mixed features. The experimental results showed that the algorithm performed better in the face of mixed feature extraction. Compared with the average value of a single feature, the recall and precision of the three mixed features increased by 13.78% and 15.64%, respectively. Moreover, in the case of a large number of concurrent features, the algorithm also met the basic requirements. When the concurrent number was 100, the average response time of the algorithm is 4.46 seconds. Therefore, the algorithm proposed by the research institute effectively improves the ability of video retrieval and can meet the requirements of timeliness, which can be widely applied in practical applications.**

Keywords*—Matching extraction; feature fusion; image retrieval; intelligent transportation

I. INTRODUCTION

As the economy continues to grow, the basic needs of the people continue to develop, which is reflected in the field of transportation by the large increase in the number of private cars, which not only causes serious environmental pollution, but also provides greater pressure on traffic management. Therefore, China vigorously promotes the intelligent transportation, combining data-based information with traffic management to achieve automated detection and retrieval of surveillance images, which greatly improves the efficiency of transportation and reduces the burden of manpower, and is an effective way to retrieve target images [1]. Today's video image retrieval systems have been able to achieve real-time extraction functions, while providing appropriate processing methods. First of all, it is necessary to install surveillance at different intersections or streets, and the intelligent transportation system is divided and displayed according to the area. The supervisory equipment of the video can help the relevant staff to realize the real-time processing of effective information, and also has five functions of list classification, live video, quick screenshot, remote lens control and setting up display image parameters. Among them, the list display can make the video classification more efficient. The system automates the display of list data through a tree diagram format, with the installation location as the title and the boundaries according to each area, and can usually view 1 to 16 surveillance screens simultaneously. It can also be further assisted by features such as quick screenshots and remote control of footage. Human adjustment of each parameter can also be utilized during live video streaming to assist in completing the work efficiently. Administrator-centric maintenance of organizational units, devices, i.e., surveillance points, and electronic maps can be implemented at [2, 3]. Image retrieval can be applied to areas such as traffic flow statistics, i.e., the use of image features to count the flow of people and vehicles. These informational functions greatly reduce the pressure of traffic management, because the traditional manual detection methods are not only time-consuming and laborious, but also have a very limited scope of work. The new digital management can help traffic management to be smoother and more accurate [4]. The video image retrieval system can provide great value to the traffic field, so the study explores the intelligent traffic system based on image processing and feature extraction algorithm, and designs three stages of data pre-analysis of traffic, feature matching of video and image, and feature extraction, and the study aims to further enhance the development of intelligent traffic. The main contribution of the research is to design and build the architecture of video surveillance systems, complete the collection and monitoring of image data, and then utilize video stream keyframe extraction technology to integrate various image analysis and processing sub-functional modules, forming effective algorithms for analyzing and identifying surveillance objects.

The research content is mainly divided into four parts. The first part is a summary and analysis of domestic and foreign scholars' research on image retrieval technology and image processing technology. The second part is to study and construct a smart transportation model, which describes the extraction of image color and texture features. The third part is to conduct performance experiments on the proposed method

through experiments, and verify the feasibility of the research method through scientific control. The fourth part summarizes the research and analyzes the shortcomings in the current research, while proposing future research directions.

II. Related Works

The image retrieval function of video has been studied by many scholars in various fields. Yan et al. [5] concluded that video image retrieval also has some security privacy issues. Based on this, they proposed a new class of secure video retrieval method to maintain user privacy based on cryptographic type vector expansion approach, and experimentally verified that this cryptographic retrieval yielded the same results as ordinary retrieval. Radenović et al. [6] proposed a convolutional neural network based image retrieval method; This method is highly compact as well as extremely efficient in extraction, at the same time it also requires a large amount of data support for the training of the model; The study used the method of adjusting cellular neural networks, using automated retrieval of random images, introducing concepts such as hard positives. This is to further improve the efficiency of the algorithm and achieve efficient operation of image retrieval. Veres and Moussa [7] considered that the transportation system is a more complex. This system requires the modules to operate in cooperation with each other, which contains a large number of temporal and spatial features, so building a model for this system is difficult. The study then proposes a deep learning theory, gives a detailed overview of its development in transportation systems, and provides solutions for its development and other problems. Chen et al. [8] argued that traditional transportation systems should have been overturned long ago. Consequently, the study applies a deep learning theory based on edge nodes to traffic flow data, introduces a detection algorithm for YOLOV3 and a deep simple online tracking algorithm for vehicle detection purpose. Gohar and Nencioni [9] combined smart transportation with 5G technology. Then they provide an overview of the background as well as the prospect of 5G technology and describe its application in smart transportation system in detail and also analyze its application in other fields. The study uses image processing based as well as feature extraction algorithms to design traffic systems. Rovithakis et al. [10] used hybrid neural networks as well as genetic algorithms for feature extraction and apply it to medicine. This is to identify normal and cancerous cells whose features generate highly dispersed classes in space while using spectral classification to achieve further testing. Xu et al. [11] considered the use of the study discussed the temporal characteristics of the data and proposed a nonlinear data feature extraction technique, i.e., kernel principal component analysis. This technique first maps the low-dimensional data into a high-dimensional space and completes the feature extraction in that control. Then the study compared this method with the linear PCA method experimentally, and obtained that the kernel principal component analysis method has better. The key to more efficient feature extraction is to compress the data, which can achieve the purpose of discarding irrelevant signals, excluding noise and various redundant feature data. Zebari et al. [12] used a combination of FS and FE method for feature extraction and verified the reliability performance of the method through comparative experiments.

In the above analysis, scholars have used various algorithms to retrieve videos or images, etc., in order to achieve intelligent transportation. Their research used combinatorial algorithms to construct models and obtained relatively reliable results. However, the results basically only meet the minimum requirements for practical applications, and in the face of complex video environments, its performance is difficult to achieve the same effect. Therefore, research suggests that in smart transportation, algorithms still need to have better performance. Inspired by the above literature, the study combines image processing technology, feature extraction technology, and other technologies to form a composite model. This model is applied to the field of transportation.

III. Intelligent Traffic System Based on Video Image Retrieval

A. Intelligent Transportation System Data Pre-processing

The study adopts a video retrieval system based on image processing technology and feature extraction algorithm. The system contains five main aspects, which are: image processing, data storage, monitoring and management, and traffic flow statistics. The flow chart of the whole system is shown in Fig. 1.

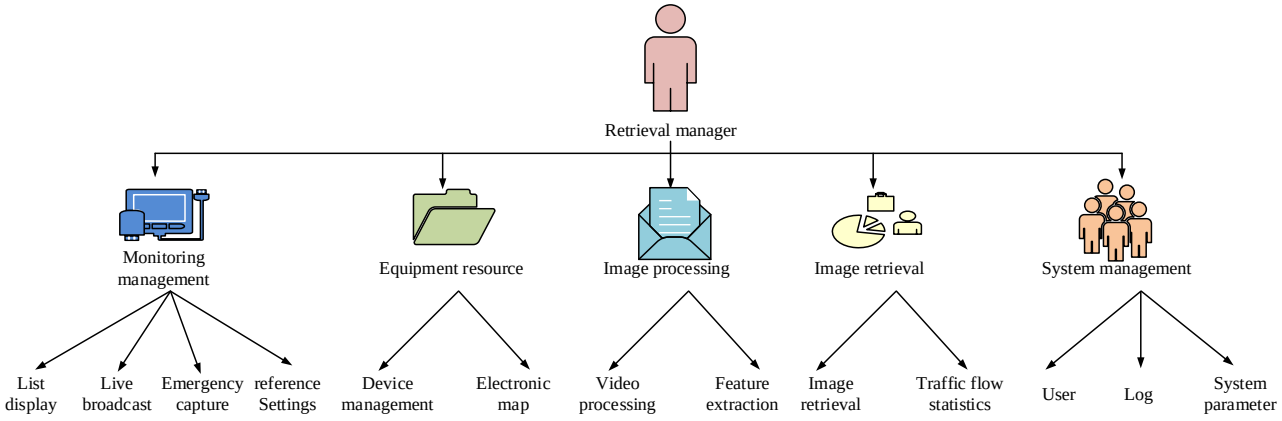


Fig. 1. Intelligent transportation system.

Data preprocessing is performed in order to eliminate as many extraneous signals as possible and to make the subsequent image processing steps more efficient. Data preprocessing includes data normalization, noise exclusion, and attenuated luminance [13]. The formula is shown in Equation (1):

$$f(D)=\frac{D_m}{A_0}\int_0^D H(u)du=D_m\int_0^D\frac{H(u)}{A_0}=D_mP(D) \quad (1)$$

In Equation (1), D_m denotes the maximum gray value in the image set; A_0 denotes the total pixel value in the image set; $H(u)$ denotes the total number of all pixels in the image set whose gray value is not equal to 0; P represents the probability of the occurrence of grayscale values in an image. This equalization method is based on the mutual difference relationship of pixel gray values for feature recognition. The difference between the gray value of a pixel and its neighboring pixels within a certain range is used as the judgment criterion. When it exceeds a certain threshold value, it means that the point is an irrelevant signal, i.e., it has a correlation crossover with other points and should be discarded; on the contrary, the point can be regarded as a valid signal to be retained, and the expression of the whole process is shown in Equation (2):

$$y_{ij}=\begin{cases} med(W[x_{ij}]), x_{ij}=\min or\max \\ x_{ij}, x_{ij}=other \end{cases} \quad (2)$$

In the above Equation (2), y_{ij} represents the pixel value of the output point (i, j) after equalization; x_{ij} represents the initial pixel value of the input; W represents the weighting operation; and med represents the intermediate value function of the solution. In summary, the pre-processing of video, i.e., screening key frames and eliminating noisy signals, finally constructs an image basic element that can realize feature extraction. The basic principle of feature matching is to realize the feature matching process by comparing the feature data of the desired image with the image feature information in the database, and using the similarity as the judgment criterion. There are five main categories of feature vector matching algorithms that are relatively advanced in development today. One of them is the histogram crossover algorithm described above, and the second one is the cardinality split-box algorithm, i.e., feature matching by means of cardinality test, when the cardinality value χ is small, the correlation is weak, and vice versa, the correlation is strong [14]. The expression is shown in Equation (3):

$$\chi^2=(Q,I)=\sum_i\frac{(Q_i-I_i)^2}{(Q_i+I_i)^2} \quad (3)$$

In the above Equation (3), Q and I represent the histogram of the target image and the histogram of any image in the database, respectively, i represents the stalk in each histogram. The Kolmogorov-Smirnov algorithm is a method of dimensionality reduction of data using spectral analysis, which can reduce the storage cost, but a large number of decomposition clusters will increase the computational pressure and lead to lower efficiency. The distance method L^p is calculated as shown in Equation (4) [15].

$$L^p(x,y)=(\sum_i|x_i-y_i|^p)^{1/p} \quad (4)$$

In the above Equation (4), x and y represent the contents of the target image and the value of the image matched with it in the database, respectively. When the parameter p is 1, the algorithm uses the Euclidean paradigm, i.e., the distance of $L1$, as the basis of the calculation result; when the parameter p is 2, the Manhat tan paradigm, i.e., the distance of $L2$, is used as the basis of the calculation result. Considering according to the research needs and combining the advantages and disadvantages performance of each algorithm, the study selected L^p distance method as the matching feature method, which is not only applicable, but also has short computation time and relatively easier to implement [16].

B. Extraction of Color Features in Video Images

The image retrieval requires a combination of feature matching and feature extraction. Then comes the feature extraction of video images, which can be roughly divided into three parts: color feature extraction, texture feature extraction, and shape feature extraction. In this study, color moments are used to achieve feature extraction, and the overall color feature matching process is shown in Fig. 2.

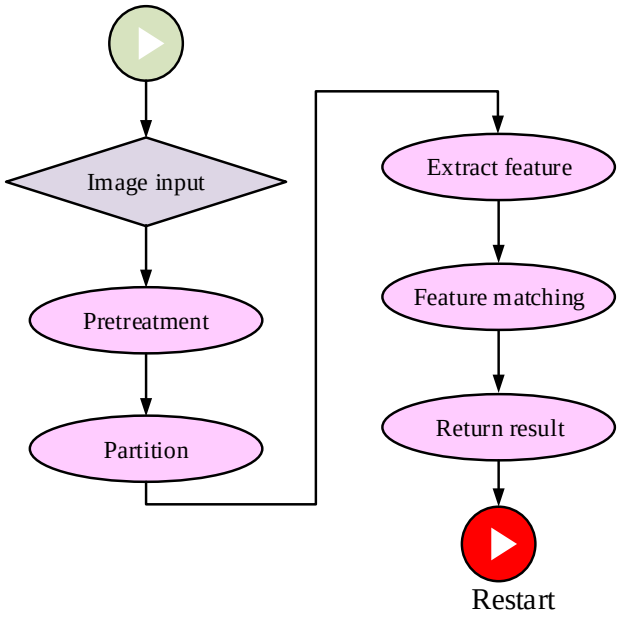


Fig. 2. Color feature extraction.

Different color models usually differ in their feature representation methods. The study is modeled by the perceptual color model HSV (Hue, Saturation, Value). This model describes colors according to the three attributes of hue, luminance and saturation, and input to the Munsell 3D spatial coordinate system for further analysis. Generally, video images are stored in the RGB model, so the first step is to change the video image from the traditional RGB format to the HSV

format [17]. It is known that all parameter values in the RGB model (r,g,b) are in the interval [0,255], and the conversion expression is shown in Equation (5):

$$\begin{cases} s = \dfrac{\max(r,g,b)-\min(r,g,b)}{\max(r,g,b)} \\ h' = \begin{cases} (5+b'), if & r=\max(r,g,b) \quad and \quad g=\min(r,g,b) \\ (1-g'), if & r=\max(r,g,b) \quad and \quad g\neq\min(r,g,b) \\ (1+r'), if & g=\max(r,g,b) \quad and \quad b=\min(r,g,b) \\ (3-b'), if & g=\max(r,g,b) \quad and \quad b\neq\min(r,g,b) \\ (3+g'), if & b=\max(r,g,b) \quad and \quad r=\min(r,g,b) \\ (5-r'), otherwise \end{cases} \\ h = 60\times h' \end{cases} \tag{5}$$

It is known that the values of h in the HSV model are in the interval [0,360], the values of s and v are in the interval [0,1], and all the parameter values of (r,g,b) in the RGB model are in the interval [0,255], when $v' = \max(r,g,b)$, then $v = \dfrac{v'}{255}$ is defined. In the above Equation (5), the expressions of r', g' and b' are shown in Equation (6):

$$\begin{cases} r' = \dfrac{v'-r}{v'-\min(r,g,b)} \\ g' = \dfrac{g'-g}{v'-\min(r,g,b)} \\ b' = \dfrac{b'-b}{v'-\min(r,g,b)} \end{cases} \tag{6}$$

The HSV color model has the most accurate description in terms of individual subjective perception. To further improve the accuracy of the model, the study segmented the output video images in the form of 3 × 3. Then it ensured that the description values of the three different dimensions were normalized and pre-processed before extracting the features, which can avoid the computational error of taking too high a local value. The accuracy of video images retrieved using only histograms is not high because histograms are weak in considering spatial location, lighting intensity and element correlation; This can lead to more subjective results [18]. Based on this, the study introduces the "color-space" theory, the principle of which is to infinitely partition the color space, and extract the corresponding color features from the divided small space, and finally form a sequence of feature vectors. The similarity of the feature vectors is used to compare and sort the video images for retrieval. This recognition method, which takes into account the spatial location information, greatly improves the accuracy of the model. In the retrieval of each feature in space, the center distance according to the three orders of the image is used as the index term, denoted by E, σ, and S, corresponding to the description of the color mean, standard variance, and cubic root asymmetry, as shown in Equation (7):

$$\begin{cases} E = \dfrac{1}{A}\sum_i\sum_j p_{ij} \\ \sigma = \left[\dfrac{1}{A}\sum_i\sum_j (p_{ij}-E)^2\right]^{\frac{1}{2}} \\ S = \left[\dfrac{1}{A}\sum_i\sum_j (p_{ij}-E)^3\right]^{\frac{1}{3}} \end{cases} \tag{7}$$

In the above Equation (7), A represents the number of pixel values of the complete image; E also represents the pixel values at the (i,j) coordinates. The image retrieval is achieved by comparing the target image with the color principal feature vector of the retrieved image $Vector(E,\sigma,S)$, and the expression is shown in Equation (8):

$$D(Q,I) = D(Vector_Q, Vector_I) = W_E\left|E_Q - E_I\right| + W_\sigma\left|\sigma_Q - \sigma_I\right| + W_S\left|S_Q - S_I\right| \tag{8}$$

In the above Equation (8), Q represents the target image; I represents an arbitrary image in the database; $D(Q,I)$ represents the similarity value between the two images; $Vector_Q$ and $Vector_I$ represent the feature vectors of the target image and the arbitrary image, respectively; W_E, W_σ, and W_S represent the weights corresponding to the color mean, standard variance, and cubic root asymmetry, respectively. Since the color mean is greatly affected by the light intensity, the study will moderately reduce its weight value.

C. Extraction of Texture Features in Video Images

The study uses a grayscale co-occurrence matrix to extract the texture features of the image, which can be used to describe the pixel changes in a certain direction in space. The position distance of two pixels is defined as $\delta = (D_x, D_x)$ and their probability of occurrence can be expressed by the gray value $P(i,j|\delta,\theta)$. Under the condition that the position δ and the space θ are certain, the gray value can be simplified to $P(i,j)$ and the parameter (i,j) takes a range of values related to the number of gray levels L, which is a natural number less than or equal to $L-1$. The grayscale co-occurrence matrix uses the grayscale correlation of each pixel in the space to predict the probability of a certain grayscale value, and ultimately to achieve the description of texture features [19]. However, the computational pressure of the model would be too high if the probability of a gray value is calculated for all locations in the space. The study introduces moment of inertia, energy, entropy, and correlation as the four basic attributes of the feature vector. The flow of the algorithm implementation is shown in Fig. 3.

From Fig. 3, the system should first convert the original image to a gray image, and the conversion formula is shown in Equation (9):

$$gray = 0.30\times R + 0.59\times G + 0.11\times B \tag{9}$$

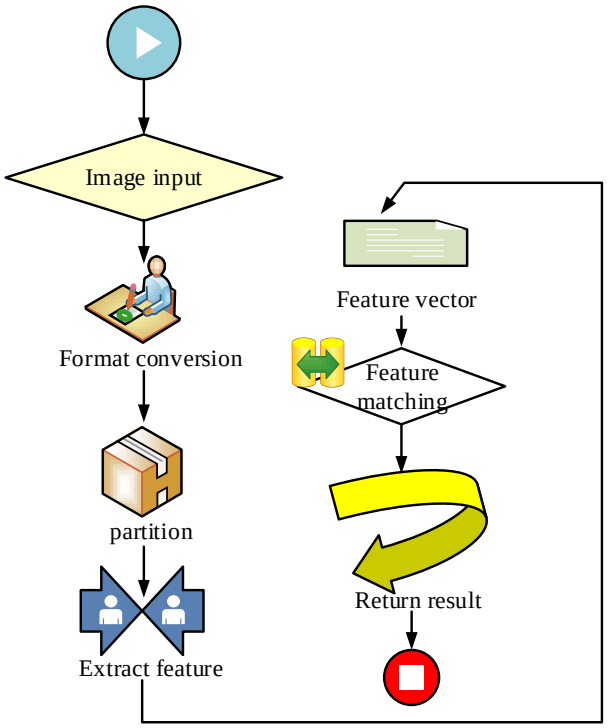


Fig. 3. Texture feature extraction.

where, the number of gray levels L is 256; however, too many gray levels will lead to the non-representation of viewing differences and increase the computational pressure. Consequently, before creating the gray symbiotic matrix, the number of gray levels should be compressed in advance, and the study sets the total number to 8, as shown in Equation (10):

$$\begin{aligned} Q1 &= \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} p(i,j)^2 \\ Q2 &= \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} (i-j)^2\, p(i,j) \\ Q3 &= \left\{ Q1 = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} (i+1)(j+1)\, p(i,j) - u_1 u_2 \right\} / (\delta 1 \delta 2) \\ Q4 &= \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} p(i,j) \log_2 p(i,j) \end{aligned} \tag{10}$$

The four matrices constructed are all characterized by the four perspectives of $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$ for the eigenvalues, and the interval between pixels d is set to 1. Then the pixels that meet the conditions such as location features are filtered and filled into the matrices respectively. The four attributes of each feature covariate in the formed gray co-occurrence matrix are further analyzed and calculated. Finally, the features are extracted and formed into a 4 × 4 matrix. The mean as well as the standard deviation of the texture features are calculated. The sequence is shown in Equation (11) as follows:

$$F = \left[\mu_1, \mu_2, \mu_3, \mu_4, \sigma_1, \sigma_2, \sigma_3, \sigma_4 \right] \tag{11}$$

It is also necessary to normalize the parameters in the above Equation (11) and rank the similarity by the Euclidean distance method to retrieve the most appropriate image.

D. Extraction of Shape Features in Video Images

Shape features are one of the most important aspects of image matching. The shape analysis starts with enclosing the region with a closed curve, which can be divided into two types: by boundary and by region. The extraction of shape features focuses on shape area, aspect ratio, and moment invariants, and therefore invariant moments are used to extract shape features. However, image area, orientation, and distortion affect the accuracy of the algorithm, so it is necessary to introduce algorithms that change with these three aspects to achieve feature extraction. The complete flow of shape feature extraction is shown in Fig. 4.

Fig. 4 showcases that the study first extracts image contours as well as core features based on the canny operator method. It is a method for boundary detection based on the first-order derivatives of Gaussian functions and strong and weak thresholds. However, there may be breakpoints at the edges, and the study uses a closed-loop operation to connect the edge breakpoints to form a complete closed loop to avoid the influence of noise signals. Then the seeds are filled using the diffuse water method, and whether they are filled or not is judged according to whether they match the pixel values of the initial image; If they are filled, the lightness and darkness of the pixel point as well as the color value need to be adjusted until all points within the closed loop are finished testing [20]. Then the invariant moments in the closed loop are calculated and the feature vectors are refined. The main purpose is to calculate the 7 HU invariant moment features in the image, and further revise the calculation results before completing the final vector sequence. Finally, a normalization operation is performed on all element values using Gaussian method to ensure that there is a difference between the elements. And even if there are extreme values, the complete shape matrix is guaranteed and normalized as shown in Equation (12):

$$\begin{cases} h_1 = \eta_{20} + \eta_{02} \\ h_2 = (\eta_{20} - \mu_{02}) + 4\eta^2{}_{11} \\ h_3 = (\eta_{30} - 3\mu_{12})^2 + (3\eta_{21} - \mu_{03})^2 \end{cases} \tag{12}$$

Finally, based on the calculation results, feature matching is performed using Euclidean distance. The calculation formula is shown in Equation (13):

$$S_{im}(Q,I) = \sqrt{\sum_{i=1}^{n} (h_i^q - h_i^l)^2} \tag{13}$$

In the above Equation (13), h_i^q and h_i^l are the normalized sequences of the two compared images. The three features of color, texture and shape of the images are considered together and the features are fused, as shown in Equation (14):

$$L^2(x,y) = \alpha \sqrt{\left(\sum_i (x_{i\alpha} - y_{i\alpha})^2\right)} + b \sqrt{\left(\sum_i (x_{ib} - y_{ib})^2\right)} + c \sqrt{\left(\sum_i (x_{ic} - y_{ic})^2\right)} \tag{14}$$

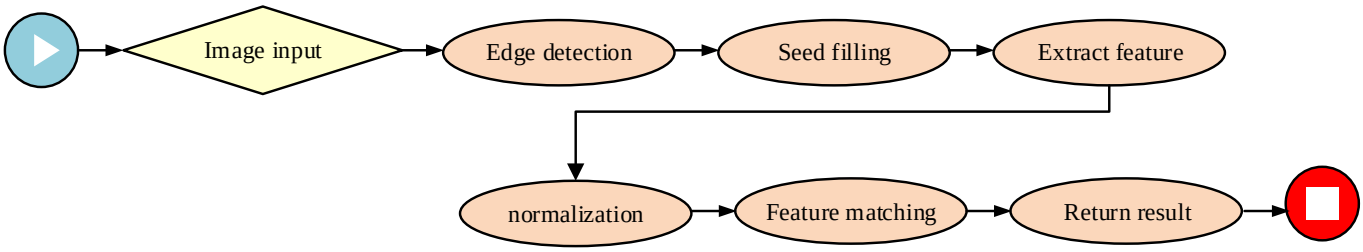


Fig. 4. Shape feature extraction.

In the above Equation (14), x and y represent the fused feature vectors of the target image and any image in the database, respectively; x_i and y_i represent the vector values of different features of each image, and abc represents the basic weights of color, texture and shape features, respectively. To ensure the normal and efficient operation of the system, the performance metrics are also required, which usually include response time, load condition and retrieval accuracy. If the bandwidth is sufficient, the response time should be less than or equal to 2 seconds, and vice versa, less than or equal to 5 seconds; And considering the possibility of a large number of search workers entering the system at the same time, the system should reach a minimum concurrency value of 100; Finally, the retrieval accuracy of the retrieval system designed in the study should be at least 75%.

IV. Simulation Experiments

The study conducts simulation experiments for the designed video image retrieval system, which can be divided into function-based experiments as well as performance-based experiments, that is, to verify the reliability of the system's business capabilities and performance indicators. The experimental environment includes both the user platform and the server platform. The specific environment parameters and data sets are shown in Table I.

The implementation of the algorithm not only requires the aforementioned hardware facilities, but also the control of algorithm-related parameters. The color value in the color space must meet the condition of [0,255], and the h parameter value converted to HSV color space should be in the range of [0,360], and the s and h parameters should be in the range of [0,1]. In the texture features, setting different concurrency numbers and testing the response speed of the model under different conditions can obtain the results shown in Fig. 5.

Fig. 5 demonstrates that the response times from 20-100 concurrency number all meet the requirements, i.e., the response time is no more than 2 seconds when the bandwidth is sufficient, and no more than 5 seconds when the bandwidth is not sufficient. In the case where the number of concurrency is 40 or less, the model responds quickly and does not fluctuate much over multiple attempts, and is relatively stable overall. When the number of concurrency is 20, the average response time of the model is only 1.38 s. As the number of concurrency increases, the response time of the model keeps rising, but all of them are at normal values. At 60 concurrency, the average response time of the model reaches 2.75 s, and the convergence speed decreases slightly. When the number of concurrency reaches 80 and 100, the average response time of the model is 3.87 s and 4.46 s, respectively, and the fluctuation increases, but it still falls within the good convergence speed. In summary, it can be concluded that the image retrieval model is in compliance with the corresponding performance and criteria. To further understand the recognition accuracy of the model, the study conducted a comparative test of the model according to different feature extraction requirements, which are known to be 40 for all retrieval results and 35 for all suspicious images, and the results of the number of recognition when only a single feature is retrieved are shown in Fig. 6.

TABLE I. Environmental Parameters and Data Sets

Database/server platform					
Configuration Content	CPU	Memory	Hard drive capacity	Operating System	Database/ Development Environment
Configuration details	5.4GH. Intel i5. Quad-Core Processor	16GB	20TB/ 5TB	Windows Server 2012.	Oracle1ln Eclipse112
User Platform					
Configuration Content	Model		Operating system		Quantity
Configuration details	DELL530, Intelis, Memory 8G.		Windows 7		10
Number	1	2	3	4	5
Frame Rate	1525	1841	1562	1845	1956
Detailed parameters	Regular video rate, resolution 720P.				
	420 keyframes	510 keyframes	432 keyframes	395 keyframes	448 keyframes

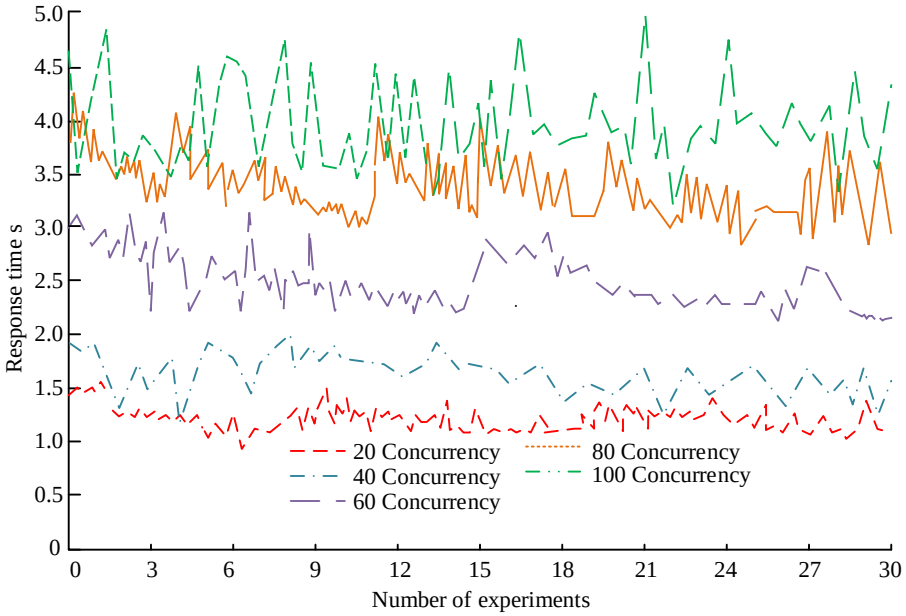
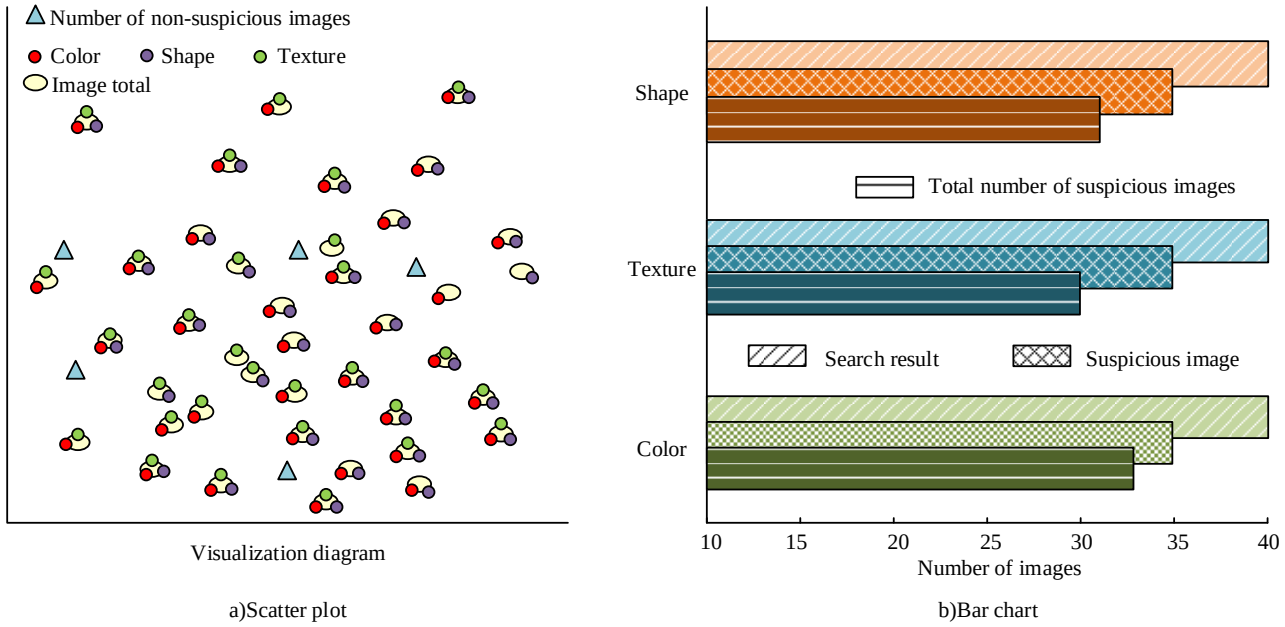


Fig. 5. Response time of the model with different number of concurrency.



a)Scatter plot b)Bar chart

Fig. 6. Quantitative result graph for single feature retrieval.

Fig. 6 showcases that there is little difference in the number of results when the model retrieves a single feature, with the known number of all images being 40, the number of suspect images being 35, and the number of retrieved color, texture, and shape features is 32, 30, and 31, respectively. The algorithm performs best when retrieving color features, due to the fact that in traffic flow, color features are usually large in area and therefore they are also the easiest to detect. Texture features, on the other hand, are often hidden deeper and are not as easy to detect. Even so, the algorithm's detection results for each feature are relatively complete. The study conducted several experiments and finally the algorithm's detection rate and accuracy on a single feature are shown in Fig. 7.

Fig. 7 showcases that the retrieval accuracy of the algorithm for a single feature is in the reasonable range, with the highest retrieval accuracy and recall for a single color feature, 77.34% and 83.21%, respectively. Compared with the single texture feature with the worst retrieval performance, it has improved by 1.47% and 7.06%, respectively. It demonstrates that the accuracy of the algorithm is closely related to the feature attributes. Large area features are more easily recognized by the algorithm, and the average recall and precision of the algorithm on a single feature are 76.45% and 79.32%, respectively. Then, further analysis was conducted on the retrieval of mixed features in the experiment, and the results shown in Fig. 8 were obtained.

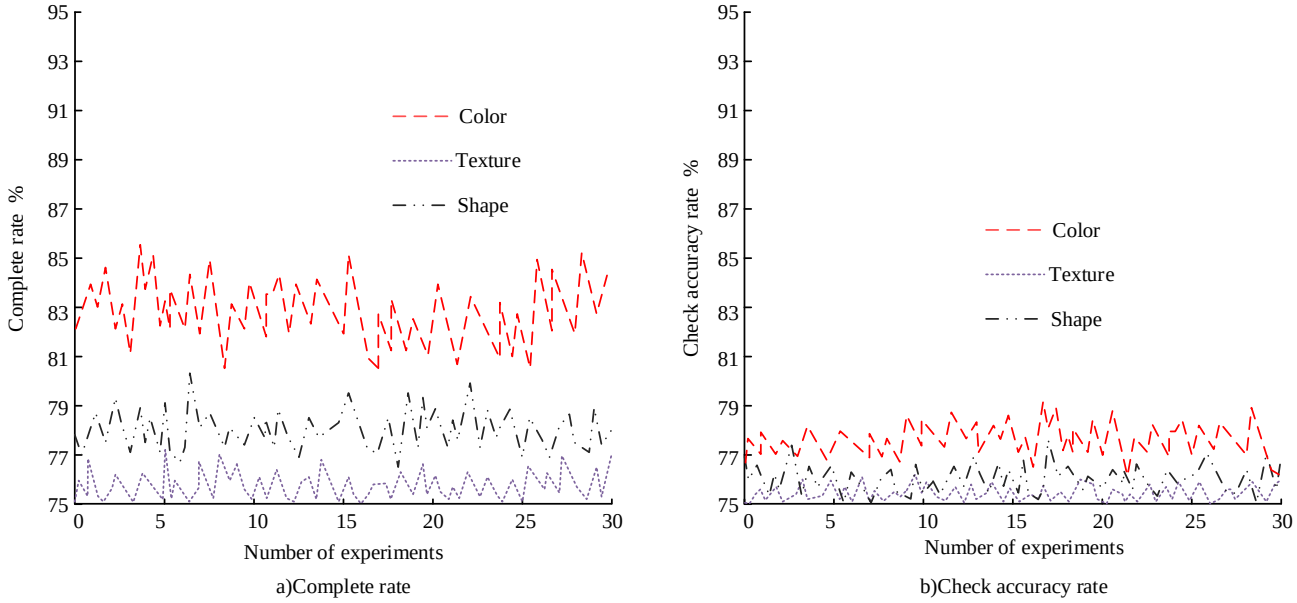


Fig. 7. Retrieval accuracy of single feature.

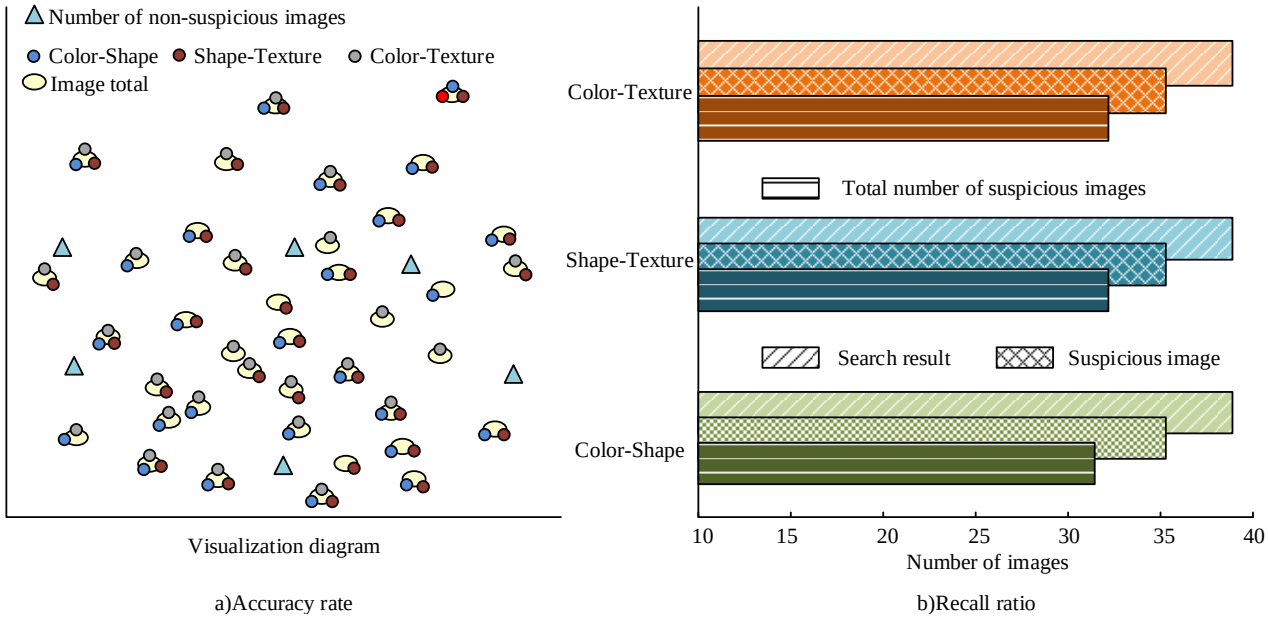
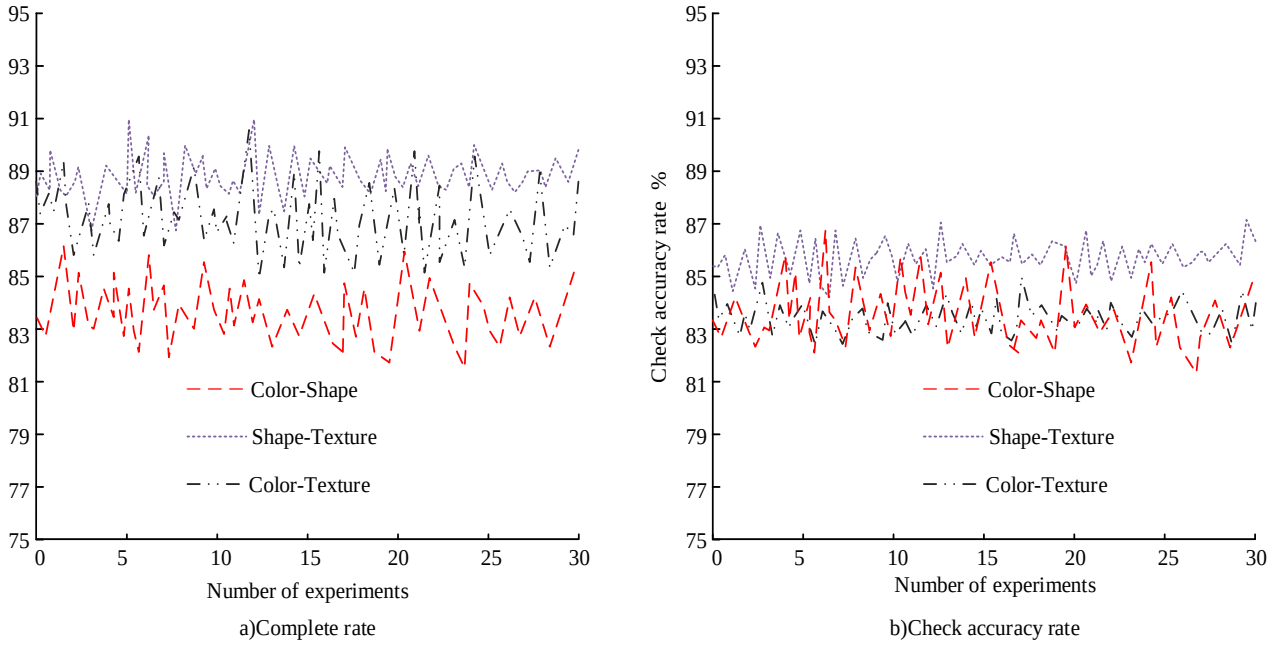


Fig. 8. Number of correct retrievals of mixed features.

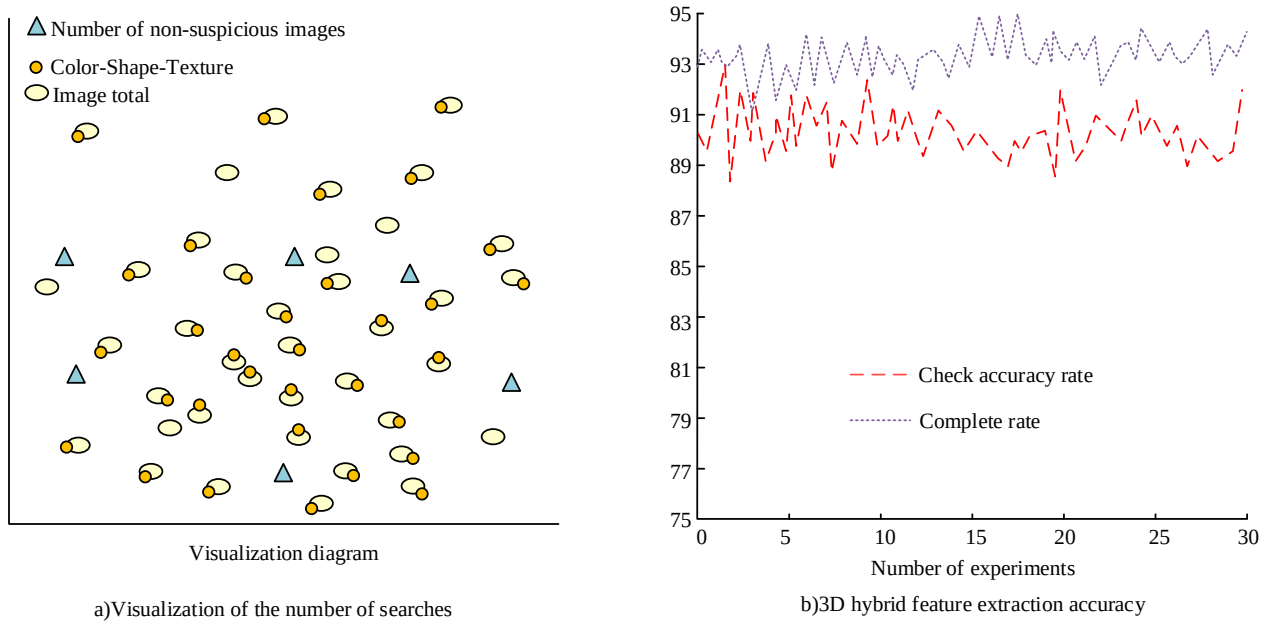
Fig. 8 demonstrates that the model has a corresponding improvement for the retrieval of mixed features, which are color-shape, color-texture, and shape-texture, respectively. There are some differences in the correct retrieval values, but they can be ignored. The correct retrieval values of the three mixed features are 32, 33, and 33, which are less different from the correct values. The model's retrieval values for the color-shape hybrid features are relatively weak, but are within the good range. Thus the model also has excellent performance for the retrieval of both hybrid features. The study conducts several experiments and finally results in the algorithm's find-all and find-accurate rates on the two mixed features as shown in Fig. 9.

Fig. 9 showcases that the algorithm has significantly improved its performance in mixed features compared to single features. Among the three mixed features, the shape texture retrieval ability is the best, with a precision and recall rate of 85.35% and 88.24%, respectively. Compared with color shape features, the algorithm has improved by 1.13% and 3.45%, respectively. The average retrieval recall and precision of two-dimensional mixed features were 85.37% and 87.05%, respectively, which increased by 8.92% and 7.73% compared to single features. The study further validated the algorithm for 3D mixed features and obtained the results shown in Fig. 10.



a)Complete rate b)Check accuracy rate

Fig. 9. Retrieval accuracy of two-dimensional features.



a)Visualization of the number of searches b)3D hybrid feature extraction accuracy

Fig. 10. Retrieval accuracy of 3D features.

Fig. 10 illustrates that the increase of feature parameters is not a decrease of algorithm accuracy, but an effective improvement. The total number of detected suspicious images reaches 34, which is only one bit away from the true value, and the detection accuracy and completion rates reach 90.23% and 94.96%, respectively. Compared with the retrieval averages of single features, the improvement is 13.78% and 15.64%, respectively; compared with the retrieval averages of 2D mixed features, the improvement is 4.86% and 7.91%, respectively. It can be seen that the increase of feature dimensionality leads to higher accuracy of the algorithm. In summary, it can be concluded that the accuracy of the model meets the corresponding criteria and is excellent in the performance of mixed features.

To further validate the feasibility of the proposed method, the study conducted comparative experiments with other advanced methods to verify the feasibility of the research method. The experiment used Adaboost algorithm, Mean Shift algorithm, and Hard Triplet algorithm for comparative experiments. The model performance was tested through detection accuracy and recall indicators, and the results are shown in Table II.

TABLE II. ALGORITHM PERFORMANCE COMPARISON RESULTS

Algorithm	Rrecision	Recall	Reaction time (s)
Research algorithms	0.964	0.951	0.62
AdaBoost algorithm	0.958	0.946	0.58
Mean shift algorithm	0.967	0.948	0.64
Hard Triplet algorithm	0.955	0.952	0.59

In Table II, the accuracy of the AdaBoost algorithm is 0.958, the recall rate is 0.951, and the model response time is 0.58 seconds. The accuracy of the Mean shift algorithm is 0.967, the recall rate is 0.946, and the model response time is 0.64 seconds. The accuracy of the Hard Triplet algorithm is 0.955, the recall rate is 0.952, and the model response time is 0.59 seconds. It can be seen that the currently used algorithms have high accuracy and recall in video image detection, as well as fast response speed, which can meet the accuracy and real-time requirements of practical applications. The accuracy of the research method is 0.964, the recall rate is 0.951, and the reaction time is 0.62 seconds. From the analysis of the results, the proposed method has a relatively close performance to the current advanced methods, with an accuracy second only to the Mean shift algorithm, a recall second only to the Hard Triplet algorithm, and a reaction time difference of only 0.4 seconds. Therefore, the method proposed in the study can meet practical needs in use.

V. CONCLUSION

The intelligent traffic system integrates digital information technology into traffic management, and is a novel method to achieve real-time and efficient monitoring and retrieval of traffic flow. The study has developed and designed the video image retrieval module in Java EE platform. Pre-processing measures such as normalization and equalization are performed to reduce the influence of irrelevant signals. L^p distance method and Euclidean distance method are used for data feature fusion and key frame extraction, and finally feature extraction methods are introduced according to three features: color, texture and shape, respectively. The study conducted simulation experiments on the model, and firstly, the convergence performance of the algorithm was tested, divided into 20-100 concurrency experiments. The experimental results show that the convergence effect of the algorithm decreases as the number of concurrency increases, but they are within the appropriate range, and when the number of concurrency is 20, the model retrieval time is only 1.38 s; and when the number of concurrency reaches 100, the retrieval time also reaches 4.46 s, which is less than the allowed the maximum value. Then the model was experimented for feature extraction, and several trials were done according to 1D, 2D and 3D features. The experimental results showed that the more feature elements, the higher the accuracy of the algorithm. The retrieval accuracy and completeness of 3D features, compared with the retrieval average of single features, improved by 13.78% and 15.64%, respectively. It can be seen that the model has better retrieval performance. However, the model also has some limitations, because video image retrieval should have a place in both the intelligent transportation field and other fields. This requires that the system should be extensible so that new modules can be added at any time. The study has yet to improve its performance in this regard.

REFERENCES

[1] S. Yadav and R. Rishi, "Secure and authenticate communication by using SoftSIM for intelligent transportation system in smart cities," J. Phys.: Conf. Ser., vol. 1767, no. 1, pp. 12049-12050, Feb, 2021.

[2] A. Sumalee and H. W. Ho, "Smarter and more connected: Future intelligent transportation system," IATSS Res., vol. 42, no. 2, pp. 67-71, Jun, 2018.

[3] T. Hassan, A. El-Mowafy, and K. Wang, "A review of system integration and current integrity monitoring methods for positioning in intelligent transport systems," IET Intell. Transp. Sy., vol. 15, no. 1, pp. 43-60, Jan, 2021.

[4] M. Asadi, M. Fathy, H. Mahini, and A. M. Rahmaniet, "A systematic literature review of vehicle speed assistance in intelligent transportation system," IET Intell. Transp. Sy., vol. 15, no. 8, pp. 1973-1986, Aug, 2021.

[5] H. Yan, M. Chen, L. Hu, and C. Jia, "Secure video retrieval using image query on an untrusted cloud," Appl. Soft Comput., vol. 97, no. 4, pp. 106782-106782, Dec, 2020.

[6] F. Radenović, G. Tolias, and O. Chum, "Fine-tuning CNN image retrieval with no human annotation," IEEE T. Pattern Anal., vol. 41, no. 7, pp. 1655-1668, Nov, 2018.

[7] M. Veres and M. Moussa, "Deep learning for intelligent transportation systems: A survey of emerging trends," IEEE T. Intell. Transp., vol. 21, no. 8, pp. 3152-3168, Jul, 2019.

[8] C. Chen, B. Liu, S. Wan, P. Qiao, and Q. Pei, "An edge traffic flow detection scheme based on deep learning in an intelligent transportation system," IEEE T. Intell. Transp., vol. 22, no. 3, pp. 1840-1852, Oct, 2020.

[9] A. Gohar and G. Nencioni, "The role of 5G technologies in a smart city: The case for intelligent transportation system," Sustainability, vol. 13, no. 9, pp. 5188-5188, May, 2021.

[10] G. A. Rovithakis, M. Maniadakis, and M. Zervakis, "A hybrid neural network/genetic algorithm approach to optimizing feature extraction for signal classification," IEEE T. Syst. Man Cy. B, vol. 34, no. 1, pp. 695-702, Mar, 2019.

[11] D. W. Xu, Y. D. Wang, P. Peng, Y. Liu, and X. Xiao, "Kernel PCA for road traffic data nonlinear feature extraction," IET Intell. Transp. Sy., vol. 13, no. 8, pp. 1291-1298, Aug, 2019.

[12] R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari, and J. Saeed, "A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction," J. Appl. Sci. Technol. Trends, vol. 1, no. 2, pp. 56-70, May, 2020.

[13] A. Mannan, K. Javed, A. Rehman, S. K. Noon, and H. A. Babri, "Optimized segmentation and multiscale emphasized feature extraction for traffic sign detection and recognition," J. Intell. Fuzzy Syst., vol. 36, no. 1, pp. 173-188, Feb, 2019.

[14] R. Wang, W. Zhang, Y. Shi, X. Y. Wang, and W. M. Cao, "GA-ORB: A new efficient feature extraction algorithm for multispectral images based on geometric algebra," IEEE Access, vol. 7, no. 1, pp. 71235-71244, May, 2019.

[15] Y. Yang, Y. Zhou, Y. Zhang, J. Lu, H. Dong, and G. Li, "Feature extraction method of pipeline signal based on parameter optimized vocational mode decomposition and exponential entropy," T. I. Meas. Control, vol. 44, no. 1, pp. 216-231, Aug, 2022.

[16] G. Lo Sciuto, G. Capizzi, R. Shikler, and C. Napoli, "Organic solar cells defects classification by using a new feature extraction algorithm and an EBNN with an innovative pruning algorithm," Int. J. Intell. Syst., vol. 36, no. 6, pp. 2443-2464, Feb, 2021.

[17] M. Yang, "Research on vehicle automatic driving target perception technology based on improved MSRPN algorithm," J. Comput. Cogni. Eng., vol. 1, no. 3, pp. 147-151, Jan, 2022.

[18] R. J. Rajappan and T. K. Kandaswamy, "A composite framework of deep multiple view human joints feature extraction and selection strategy with hybrid adaptive sunflower optimization-whale optimization algorithm for human action recognition in video sequences," Comput. Intell., vol. 38, no. 2, pp. 366-396, Jan, 2022.

[19] A. R. Lubis, M. K. M. Nasution, O. S. Sitompul, and E. M. Zamzami, "The feature extraction for classifying words on social media with the Naïve Bayes algorithm," IAES Int. J. Artif. Intell., vol. 11, no. 3, pp. 1041-1048, Sep, 2022.

[20] X. Hu, Y. Li, L. Jia, and M. Qiu, "A novel two-stage unsupervised fault recognition framework combining feature extraction and fuzzy clustering for collaborative AIoT," IEEE T. Ind. Inform., vol. 18, no. 2, pp. 1291-1300, Apr, 2022.