IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 41, No. 5, October, 2024, pp. 2585-2592
Journal homepage: http://iieta.org/journals/ts

Optimization Strategy for Intelligent Traffic Monitoring Systems Based on Image Recognition

Yixin Ren

Faculty of International Tourism and Management, City University of Macau, Macau 999078, Macau

Corresponding Author Email: T23091105486@cityu.edu.mo

https://doi.org/10.18280/ts.410531

Received: 17 April 2024
Revised: 30 August 2024
Accepted: 21 September 2024
Available online: 31 October 2024

Keywords:
intelligent traffic monitoring, image recognition, vehicle behavior recognition, spatiotemporal features, multi-core support vector machine (SVM)

ABSTRACT

With the rapid urbanization, traffic congestion and safety issues have become increasingly prominent, and traditional traffic monitoring systems are struggling to meet the demands of modern traffic management. Intelligent traffic monitoring systems based on image recognition can significantly enhance the efficiency and safety of traffic management by analyzing traffic surveillance video data in real-time. While research in vehicle behavior recognition using intelligent traffic monitoring systems has made some progress, issues such as insufficient recognition accuracy, real-time performance, and challenges in multi-target detection and tracking remain. To address these problems, this paper proposes an intelligent traffic monitoring image-based vehicle behavior recognition algorithm. First, the overall framework of the algorithm is presented. Next, it provides a detailed introduction on constructing spatiotemporal feature bodies and analyzing the spatiotemporal feature representations of traffic surveillance videos. It then discusses the classification algorithm based on multi-core support vector machines (SVM). Finally, the algorithm's effectiveness and superiority are verified and analyzed through experimental results. This study not only enriches the theoretical framework of intelligent traffic monitoring technologies but also holds significant practical value for widespread application.

1. INTRODUCTION

With the acceleration of urbanization, traffic congestion and traffic safety issues have become increasingly severe. Traditional traffic monitoring systems are difficult to meet the growing complexity of traffic management needs, and intelligent traffic monitoring systems have emerged [1-5]. Intelligent traffic monitoring systems based on image recognition can effectively improve the efficiency and safety of traffic management through real-time analysis of traffic monitoring video data, becoming an important component of smart city construction [6-9].

Research on intelligent traffic monitoring systems has important practical significance and social value. On one hand, it can improve road traffic safety by monitoring and analyzing vehicle behavior in real time, enabling timely identification and prevention of traffic accidents. On the other hand, intelligent traffic monitoring systems can optimize traffic flow management, reduce traffic congestion, improve urban traffic efficiency, and provide more convenient services for public transportation [10-15].

Although there has been some progress in image recognition and traffic monitoring, there are still many shortcomings. Traditional methods have low recognition accuracy and real-time performance when dealing with various interference factors in complex traffic environments [16, 17]. Furthermore, existing algorithms perform poorly in multi-target detection and tracking, unable to effectively cope with challenges such as heavy traffic flow and diversified target vehicles. Therefore, there is an urgent need for a more efficient, accurate, and real-time intelligent traffic monitoring algorithm [18, 19].

This paper proposes an intelligent traffic monitoring image-based vehicle behavior recognition algorithm to address the above issues. First, the overall framework of the algorithm is presented. Next, the method of constructing spatiotemporal feature bodies and analyzing the spatiotemporal feature representations of traffic monitoring videos is detailed. It then elaborates on the classification algorithm based on multi-core SVM. Finally, the algorithm's effectiveness and superiority are verified and analyzed through experimental results. This study not only enriches the theoretical framework of intelligent traffic monitoring technologies but also holds significant practical value for widespread application.

2. INTELLIGENT TRAFFIC MONITORING IMAGE-BASED VEHICLE BEHAVIOR RECOGNITION ALGORITHM FRAMEWORK

Current intelligent traffic systems still face many challenges in complex traffic environments, such as high traffic volume, diverse vehicle types, varying environmental lighting, and occlusion issues. With the rapid development of artificial intelligence and machine learning technologies, particularly the breakthroughs of deep learning in image recognition,

intelligent traffic monitoring systems have gained more efficient technical support. By using deep learning technologies, it is possible to better extract and analyze spatiotemporal features in traffic monitoring video, improving the accuracy and real-time performance of vehicle behavior recognition.

In modern intelligent traffic monitoring systems, the accuracy and efficiency of vehicle behavior recognition directly affect the effectiveness and safety of traffic management. Different vehicle behaviors not only manifest as differences in motion trajectories but also contain deeper semantic differences. Relying solely on motion information may not be sufficient to accurately distinguish between behaviors with similar motion patterns but entirely different meanings. For example, sudden braking and normal stopping may appear similar in trajectories but have entirely different implications for traffic management. Furthermore, intelligent traffic monitoring systems have extremely high demands for the real-time and accuracy of recognition algorithms. Although many existing algorithms perform excellently on specific datasets, they often face a decrease in recognition accuracy in real-world applications under complex environments. To address these shortcomings, this paper proposes an algorithm that integrates dense trajectories with deep visual features for traffic monitoring video recognition in intelligent traffic monitoring systems. Figure 1 shows the proposed traffic monitoring video recognition algorithm framework.

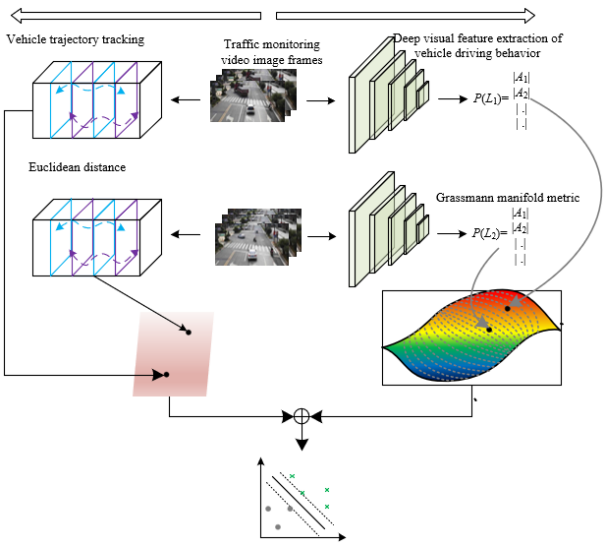


Figure 1. Proposed traffic monitoring video recognition algorithm framework

Step 1: Construct Spatiotemporal Feature Bodies

This step aims to comprehensively utilize both spatial and temporal information to provide rich feature support for subsequent behavior analysis. For each traffic monitoring video segment, the system segments it into several continuous frame images, normalizing these images to a uniform size to standardize the input data format, ensuring consistency in subsequent processing, and eliminating the impact caused by different cameras or resolution differences. Then, a convolutional neural network (CNN) is used to extract spatial features from each frame image. These spatial features construct the visual representation of each frame, which serves as the foundation for understanding the content of traffic monitoring videos. In order to retain the temporal information of traffic monitoring videos, the system sequentially arranges the extracted spatial features in chronological order, forming a spatiotemporal feature body. This ordered arrangement not only reflects the dynamic changes of vehicle behaviors but also captures the continuity of behavior occurrences. To enhance the richness of feature representation, the system samples frame images densely in a multi-scale space and performs trajectory tracking. Through trajectory tracking, the system captures the vehicle's motion trajectory in the traffic monitoring video, obtaining a large amount of action information. This motion information is crucial for recognizing specific vehicle behaviors such as sudden braking, lane changing, overtaking, etc.

Step 2: Analyze the Spatiotemporal Feature Representations of Traffic Monitoring Videos

This step aims to further process and optimize the joint features extracted from traffic monitoring videos to more accurately perform vehicle behavior recognition. For trajectory features representing temporal characteristics, these features manifest as continuous motion trajectories in traffic monitoring video sequences, and the distance between each pair of samples can be calculated in Euclidean linear space. By performing distance calculations in Euclidean space, the dynamic changes of vehicles in the time dimension, including speed, direction, and other key behavioral information, can be effectively captured, providing important evidence for recognizing vehicle movement patterns. For deep visual features representing spatial characteristics, these features, extracted by CNN, contain rich information about the vehicle's appearance and environmental background. However, directly using these high-dimensional features for calculation may lead to excessive computational complexity. Therefore, this paper adopts singular value decomposition (SVD) to reduce the dimensionality of deep visual features and extract their principal components. Since these principal components no longer reside in linear space but in a special Riemannian manifold space—the Grassmannian manifold—it is necessary to measure the distance between sample pairs in the Grassmannian manifold. Measuring distance in the Grassmannian manifold can more accurately reflect the similarity between deep visual features, thus better distinguishing different vehicle behaviors.

Step 3: Multi-core SVM Classification

This step aims to classify the spatiotemporal feature representations extracted and analyzed in the previous steps to ultimately identify and distinguish different vehicle behaviors. First, the measurement information from trajectory features in Euclidean space and deep visual features in the Grassmannian manifold is combined through linear fusion. Specifically, through the previous two steps, we obtain feature distance measurements in two different spaces: one is the trajectory feature distance based on Euclidean space, and the other is the deep visual feature distance based on the Grassmannian manifold. To effectively fuse this measurement information, this paper designs a new kernel function that linearly combines the two distance metrics, ensuring that the advantages of each feature are preserved while improving the overall recognition ability through complementary information. Then, using the newly designed kernel function, this paper employs a SVM for classification. Multi-core SVM is a powerful classifier that can handle high-dimensional and nonlinear data. By using the designed multi-core kernel function in the training phase, SVM can learn an optimal hyperplane that maximizes the separation between different categories of vehicle behaviors.

In the testing phase, the same kernel function is used to classify new samples, thereby identifying the vehicle behavior category. The advantage of SVM lies in its strong generalization ability and robustness to noise, making it particularly suitable for complex traffic monitoring environments. Through this approach, the algorithm can accurately identify and classify various vehicle behaviors, such as acceleration, deceleration, lane changing, sharp turning, etc., thus providing reliable vehicle behavior recognition results for intelligent traffic monitoring systems.

3. CONSTRUCTION OF VEHICLE BEHAVIOR SPATIOTEMPORAL FEATURE REPRESENTATION

For the traffic monitoring video set $N=\{L_1, L_2, ..., L_v\}$, this paper processes each traffic monitoring video frame by frame. Suppose the u-th traffic monitoring video L_u contains V_u frames of extracted images. This paper uses a pre-trained CNN to extract features from each image frame. In the implementation, a high-performance CNN model is selected, and the high-dimensional feature representation of each frame image is extracted from its final fully connected layer. These feature representations not only contain rich spatial information but also preserve both local and global information of the image, laying the foundation for subsequent temporal feature modeling. Once the feature extraction for each frame image is complete, the features are arranged in chronological order to form a feature sequence. This feature sequence reflects the vehicle's state at different time points in the traffic monitoring video and retains spatial detail changes of the vehicle. Next, to integrate these temporal features, this paper uses a Long Short-Term Memory (LSTM) network to process the feature sequence. By inputting the extracted CNN feature sequence into the LSTM network, the LSTM can learn vehicle behavior patterns, such as acceleration, deceleration, lane changing, etc., from the continuous frame features. Through these steps, this paper constructs a spatiotemporal feature body that integrates both spatial and temporal information.

$$D(L_u)=\left(A_1^u; A_2^u; A_3^u; ...; A_{V_u}^u\right) \tag{1}$$

This spatiotemporal feature body not only reflects the static features of the vehicle in a single frame image but also captures the dynamic changes in vehicle behavior, providing a comprehensive description of vehicle behavior.

4. VEHICLE BEHAVIOR SPATIOTEMPORAL FEATURE ANALYSIS

For the spatial feature representation of traffic monitoring videos, this paper uses a CNN to extract spatial features from each frame image and projects these features into the Grassmann manifold space for distance measurement. Specifically, suppose that for a traffic monitoring video L_u, its spatial sequence feature is represented as $D(L_u)=(A^u{}_1; A^u{}_2; A^u{}_3; ...; A^u{}_{Vu})$, where $A^u{}_k$ is the feature vector of the k-th frame image. Figure 2 shows the modeling effect of vehicle feature keypoints. To measure the distance between these high-dimensional feature vectors, this paper projects them into the Grassmann manifold space and utilizes the geometric properties of this space to capture the similarity and difference between features. This not only retains the global structural information of the features but also effectively reduces computational complexity. For the dynamic information in the traffic monitoring video, i.e., the vehicle's motion trajectory, this paper uses a dense sampling method to extract trajectory features and directly maps this trajectory information to Euclidean space for distance calculation.

For the spatial feature sequence, since each frame image's feature vector is independently and identically distributed, this paper models it using an autoregressive moving average (ARMA) model. The ARMA model effectively captures the causal relationships and state noise within the spatial sequence. Specifically, for a traffic monitoring video L_u, each feature vector $A^u{}_k$ in its spatial feature sequence $D(L_u)$ is a current state vector with state noise. This paper constructs an ARMA model to represent the linear combination of these feature vectors as a state process function $c(a_u)$. Thus, assuming that the observation matrix is represented by Z, the transformation matrix by S, and the initial state vector by $c(a_0)$, with state and observation noise both following normal distributions, i.e., $\psi \sim V(0, O)$, $\varphi \sim V(0, W)$, each segment of traffic monitoring video is associated with an ARMA model:

$$\begin{cases} d(a_u)=Zc(a_u)+\psi(u) \\ c(a_{u+1})=Sc(a_u)+\varphi(u) \end{cases} \tag{2}$$

Although maximum likelihood estimation is typically used to solve ARMA model parameters, traditional methods are not applicable in this paper due to the high-dimensional nature of the feature vectors. Therefore, this paper uses SVD to perform dimensionality reduction on the feature matrix and extract principal components to obtain the closed-form solution for the ARMA model. Specifically, for a traffic monitoring video L_u, $D(L_u)=(A^u{}_1; A^u{}_2; A^u{}_3; ...; A^u{}_{Vu})$ is the deep visual feature matrix for that video, and the implicit state estimate of the matrix is represented by $C^{\wedge v}{}_1=[c^{\wedge}(a_1), ..., c^{\wedge}(a_v)]$, $C^{\wedge v-1}{}_1=[c^{\wedge}(a_1), ..., c^{\wedge}(a_{v-1})]$, $C^{\wedge v}{}_2=[c^{\wedge}(a_2), ..., c^{\wedge}(a_v)]$. The generalized inverse of $C^{\wedge v-1}{}_1$ is represented by $(C^{\wedge v-1}{}_1)^{\dagger}$, giving:

$$\begin{aligned} \hat{Z} &= I \\ \hat{Z}_1^v &= \sum N^S \\ \hat{S} &= \hat{C}_2^v\left(\hat{C}_1^{v-1}\right)^+ \\ \hat{W} &= \frac{1}{v-1}\sum_{u=1}^{v-1}\hat{i}_u\hat{i}_u^S \end{aligned} \tag{3}$$

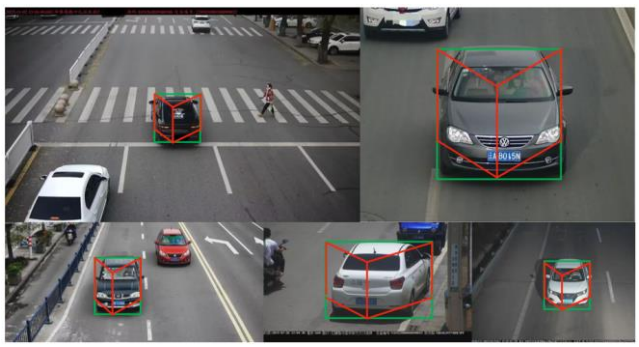


Figure 2. Vehicle feature keypoint modeling effect

In an intelligent traffic monitoring system, the core goal of the vehicle behavior recognition algorithm is to accurately identify and analyze the vehicle's behavior patterns in different traffic scenarios. To achieve this goal, the algorithm proposed in this paper utilizes the metric properties of the Grassmann manifold to effectively calculate the geodesic distance between sample pairs. Specifically, the vehicle's behavioral features can be represented as points in a high-dimensional space using an orthogonal basis, and these points form a structured space on the Grassmann manifold. Assuming the principal components of the angular vectors are represented by $\Phi=[\phi_1, \phi_2, ..., \phi_o]$, the geodesic distance calculation formula between connection points A and B on the Grassmann manifold is given by:

$$f_H(A,B)=\|\Phi\|_D \tag{4}$$

Assuming that after SVD, the principal components of the u traffic monitoring videos are represented by $I(L_u)$, the geodesic distance between deep visual features can be calculated by the following formula:

$$\begin{aligned} F_H(L_u,L_k)&=\left\|U-I(L_u)^T I(L_k)\right\|_F \\ &+\left\|U-I(L_u)^T I(L_k)\right\|_F \end{aligned} \tag{5}$$

The Frobenius norm of the matrix is denoted by $\|.\|_F$, and for each element X_{uk} in matrix X, its Frobenius norm is defined as:

$$\|X\|_F=\left(\sum_{u=1}^{l}\sum_{k=1}^{v}|X_{uk}|^2\right)^{1/2}=\left[\mathrm{tr}\left(X^H X\right)\right]^{1/2} \tag{6}$$

For the temporal motion features of the traffic monitoring video, denoted as $O(L)=(O_s, O_{s+1}, O_{s+2},...)$, this paper directly calculates the distance between sample pairs $O(L_u)$ and $O(L_k)$ in Euclidean space. The calculation formula is:

$$F_R(L_u,L_k)=\sqrt{\left(O(L_u)-O(L_k)\right)\left(O(L_k)-O(L_u)\right)^S} \tag{7}$$

5. VEHICLE BEHAVIOR CLASSIFICATION

In the intelligent traffic monitoring image vehicle behavior recognition algorithm, the choice and optimization of the classifier are key to achieving efficient and accurate recognition. First, the algorithm needs to balance the trade-off between fitting ability and generalization ability. The intelligent traffic monitoring system faces dynamic and changing traffic scenarios. While the training data is rich, the test data environment may differ significantly. Therefore, the classifier must perform well on the training samples while also possessing strong generalization ability to ensure effectiveness across various vehicle behaviors in different real-world scenarios. Second, the choice of classifier must consider the complexity of the classification function and the size of the training data. The data volume in intelligent traffic systems is vast and diverse, and the classifier must be capable of handling large-scale data while automatically adjusting its complexity. Third, the dimensionality of the feature space is another factor that requires attention. Although a high-dimensional feature space can provide more information, it may also introduce noise and redundancy, which can affect classification performance. Fourth, the homogeneity and relationships between the input feature vectors are also important considerations. For the complexity of vehicle behaviors in intelligent traffic monitoring systems, feature vectors may include various types of information, such as vehicle speed, trajectory, and position. These features may have complex interrelationships, so it is crucial to select a classifier that can capture these relationships effectively. Figure 3 gives a diagram showing the principle of vehicle behavior classification, and the basic form of the SVM is as follows:

$$\underset{q,y}{MIN}\frac{1}{2}\|q\|^2 \; s.t.\, b_u\left(q^S a_u+y\right)\geq 1, u=1,2,...,l. \tag{8}$$

By using the Lagrange multiplier method, its "dual problem" can be obtained:

$$\underset{\beta}{MAX}\sum_{u=1}^{l}\beta_u-\frac{1}{2}\sum_{u=1}^{l}\sum_{k=1}^{l}\beta_u\beta_k b_u b_k a_u^T a_k \tag{9}$$

From the above equation, the solution for β_u is the Lagrange multiplier corresponding to the training sample (a_u, b_u).

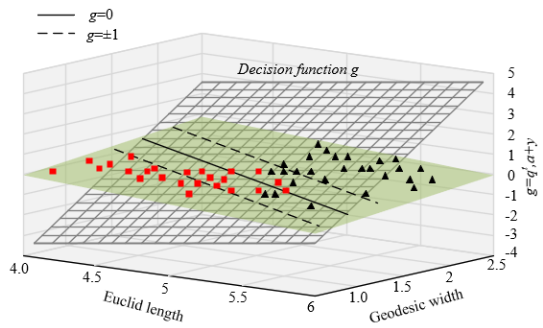


Figure 3. Principle of vehicle behavior classification

Vehicle behavior recognition in intelligent traffic monitoring systems requires handling large, complex, and nonlinear feature data. The SVM is renowned for its excellent ability to handle nonlinearity, especially when combined with a kernel function, which can effectively address nonlinear classification problems in high-dimensional feature spaces. In addition, this paper utilizes the Grassmann kernel function based on the geodesic distance between deep visual features. This kernel function is symmetric positive definite and applies to all points on the Grassmann manifold. The Grassmann manifold provides an effective way to handle data with complex geometric structures, and the geodesic distance can capture the intrinsic relationships between deep visual features. Specifically, according to Eqs. (5) and (7), the kernel function k: $H_{v,\ o}\ H_{v,o}\rightarrow E^+$ is applicable to all points on this manifold A. Therefore:

$$F_H(A_u,A_k)=F_H(A_k,A_u) \tag{10}$$

and

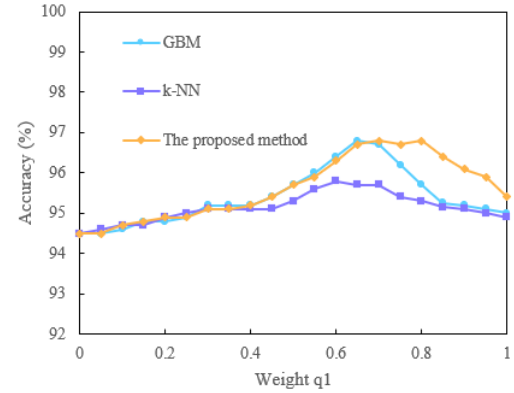
$$F_R(A_u, A_k) = F_R(A_k, A_u) \tag{11}$$

Intelligent traffic monitoring systems need to handle complex and diverse vehicle behaviors. The features of these behaviors include both temporal information, such as speed changes and acceleration, and spatial information, such as position and direction. A single feature measurement may not comprehensively capture these complex behavior patterns. Therefore, this paper opts for a multi-core SVM for training and testing. Through the multi-core learning approach, different kernel functions can handle different types of features. For example, one kernel function can process the temporal features of the video, while another kernel function can handle spatial features. This method not only fully utilizes the advantages of various features but also automatically adjusts the weights of each kernel function during the training process, enabling the model to find the best balance between different features. This ensures the system maintains efficient classification performance across different traffic scenarios and time periods. Assume that the kernel based on the Grassmann manifold space is represented by J_H, and the kernel based on the Euclidean space is represented by J_F. The weights satisfy $q_1+q_2=1$, then the following holds:

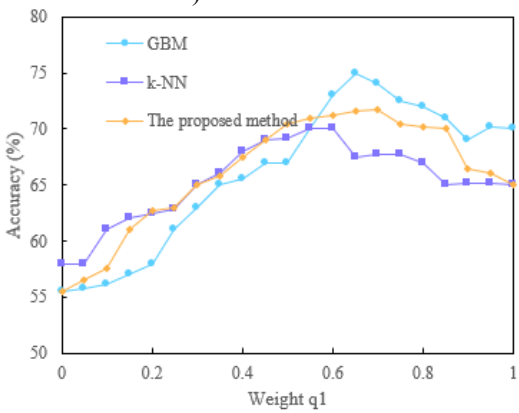
$$J(A_u, A_k) = q_1 * J_H(A_u, A_k) + q_2 * J_R(A_u, A_k) \tag{12}$$

6. EXPERIMENTAL RESULTS AND ANALYSIS

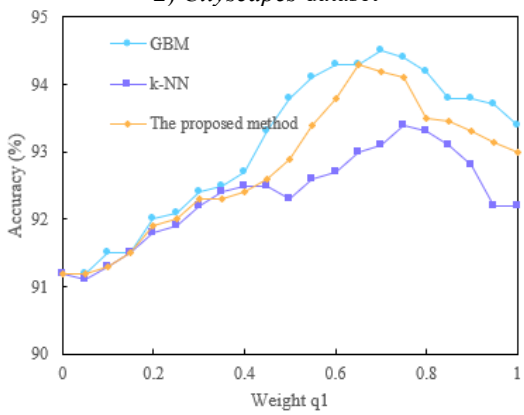
The proposed method has been validated on multiple datasets, and the experimental results shown in Figure 4 highlight its superiority. From the results on the KITTI dataset, it can be observed that the recognition accuracy of this method significantly improves as the weight $q1$ increases from 0 to 1. The accuracy rises from an initial 94.5% to a peak of 96.8%, with notable advantages at $q1$=0.8 and $q1$=1, reaching 96.8% and 96.7%, respectively. In comparison, the highest accuracy for GBM and k-NN algorithms was 96.4% and 95.8%, respectively. This indicates that the proposed method, through the fusion of spatiotemporal feature analysis, can effectively improve recognition accuracy when processing the KITTI dataset. In the Cityscapes dataset, the proposed method also demonstrated excellent performance. With an initial weight of $q1$=0, the recognition accuracy was 55.5%, and as the weight increased, the accuracy steadily improved, reaching a peak of 71.8% when $q1$=0.8. In comparison, the highest accuracy for GBM at $q1$=0.8 was 75%, while the highest accuracy for k-NN was only 70%. Although GBM slightly outperformed in some weight ranges, the proposed method performed stably and superiorly across most weight values, particularly in the medium-to-high weight ranges, demonstrating good adaptability and robustness. In the UA-DETRAC dataset, the recognition accuracy of the proposed method improved from an initial 91.2% to a peak of 94.3% (with $q1$=0.8), showing overall excellent performance. The highest accuracy for GBM and k-NN were 94.5% and 93.4%, respectively, which were very close to the proposed method. However, the proposed method exhibited more stable performance at medium and high weights and maintained a high accuracy across multiple experiments, reflecting the reliability of the algorithm. For the CCT dataset, the recognition accuracy of the proposed method increased from 83% to a peak of 91.5% (with $q1$=0.8), showing a good performance improvement trend. The highest accuracy for GBM at $q1$=0.8 was 96%, and for k-NN, it was 89%. Although GBM performed better in some weight ranges, the proposed method showed stable and reliable recognition ability overall, particularly maintaining high accuracy levels at $q1$=0.8 and $q1$=0.6.



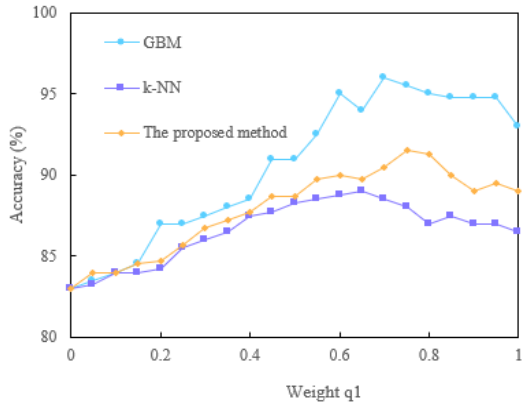
1) *KITTI* dataset



2) *Cityscapes* dataset



3) *UA-DETRAC* dataset



4) *CCT* dataset

Figure 4. Recognition accuracy of different algorithms under the weight q_1

Table 1. Experimental results of feature extraction using different dimensions of fully connected layers by different algorithms

Method / Dataset	GBM (%)		k-NN (%)		The Proposed Method (%)	
	4096 Dimensions	2056 Dimensions	4096 Dimensions	2056 Dimensions	4096 Dimensions	2056 Dimensions
KITTI	91.23	94.58	94.52	95.36	95.64	96.37
Cityscapes	68.51	71.23	70.12	73.25	71.26	80.24
UA-DETRAC	92.35	92.36	93.68	92.54	93.65	95.48
CCT	85.64	88.95	91.26	96.36	92.85	97.62

Table 2. Experimental results of different vehicle behavior spatiotemporal feature extraction models

Method / Dataset	KITTI (%)	Cityscapes (%)	UA-DETRAC (%)	CCT (%)
Proposed Method - Trajectory Features	93.5	55.8	92.3	82.4
Proposed Method - Grassmann Manifold Distance Metric	91.4	71.2	93.5	92.6
Proposed Method	94.2	88.9	94.8	94.8

To validate the effectiveness of the proposed algorithm, a comparison was made between the feature extraction results using different dimensions of fully connected layers, and a comparison with the GBM and k-NN algorithms was conducted. The detailed analysis and conclusions for each dataset are provided in Table 1.

In the KITTI dataset, the proposed method achieves a recognition accuracy of 95.64% with 4096-dimensional features, and this increases to 96.37% with 2056-dimensional features. In contrast, the GBM algorithm yields accuracy rates of 91.23% and 94.58%, respectively, while the k-NN algorithm achieves 94.52% and 95.36%. The results show that the proposed method outperforms both GBM and k-NN algorithms in both feature dimensions, especially with the 2056-dimensional features, where there is a significant improvement. This suggests that the proposed method can more accurately identify vehicle behaviors when processing the KITTI dataset by optimizing the construction and analysis of spatiotemporal features.

In the Cityscapes dataset, the proposed method's recognition accuracy with 4096-dimensional features is 71.26%, which increases to 80.24% with 2056-dimensional features. The GBM algorithm has accuracies of 68.51% and 71.23%, while the k-NN algorithm achieves 70.12% and 73.25%. The proposed method significantly outperforms other algorithms with 2056-dimensional features, demonstrating its strong feature extraction and recognition capabilities in complex urban environments.

In the UA-DETRAC dataset, the proposed method's accuracy is 93.65% with 4096-dimensional features and increases to 95.48% with 2056-dimensional features. The GBM algorithm has accuracies of 92.35% and 92.36%, and the k-NN algorithm achieves 93.68% and 92.54%. The proposed method performs excellently in both feature dimensions, particularly excelling with the 2056-dimensional features, further confirming its stability and efficiency in dynamic traffic scenarios.

In the CCT dataset, the proposed method achieves an accuracy of 92.85% with 4096-dimensional features, which increases to 97.62% with 2056-dimensional features. In comparison, the GBM algorithm reaches accuracies of 85.64% and 88.95%, while the k-NN algorithm achieves 91.26% and 96.36%. It is evident that the proposed method excels with 2056-dimensional features, significantly outperforming other algorithms and demonstrating exceptional performance in diverse traffic monitoring scenarios.

Further, a comparison was made between the performance of different spatiotemporal feature extraction models on various datasets, specifically including trajectory features, Grassmann manifold distance metric, and the proposed integrated method. The detailed analysis and conclusions for the experimental results provided in Table 2 are as follows:

In the KITTI dataset, the recognition accuracy of the proposed integrated method is the highest, reaching 94.2%. In contrast, the accuracy of the trajectory feature model and the Grassmann manifold distance metric model is 93.5% and 91.4%, respectively. Although the trajectory feature model performs similarly, the proposed integrated method still has a slight edge, demonstrating its superior ability in handling complex traffic scenarios.

In the Cityscapes dataset, the recognition accuracy of the proposed integrated method is significantly higher than that of other models, reaching 88.9%. The trajectory feature model and the Grassmann manifold distance metric model achieve accuracies of 55.8% and 71.2%, respectively. Given the more complex and variable urban traffic scenarios of the Cityscapes dataset, the proposed method's excellent performance here highlights its significant advantage in handling high-complexity scenes. By integrating multiple feature extraction methods, it can more comprehensively capture the spatiotemporal characteristics of vehicle behavior.

In the UA-DETRAC dataset, the recognition accuracy of the proposed integrated method reaches 94.8%, significantly higher than the trajectory feature model's 92.3% and the Grassmann manifold distance metric model's 93.5%. While the Grassmann manifold model performs well, the proposed integrated method improves recognition accuracy by incorporating multiple feature extraction techniques, demonstrating its strong adaptability in dynamic monitoring scenarios.

In the CCT dataset, the proposed integrated method achieves a recognition accuracy of 94.8%, significantly outperforming the trajectory feature model (82.4%) and the Grassmann manifold distance metric model (92.6%). This further confirms the outstanding performance of the integrated method in diverse and complex traffic monitoring scenarios.

In summary, through an analysis of the experimental results from different spatiotemporal feature extraction models across various datasets, the proposed vehicle behavior recognition algorithm for intelligent traffic monitoring systems shows the best overall performance. Whether in simple or complex traffic scenarios, this method can effectively improve recognition accuracy. This result indicates that the construction of spatiotemporal feature bodies and the use of multi-core SVM classification enable a more comprehensive and accurate capture of vehicle behavior features in traffic

monitoring videos, providing strong support for optimizing intelligent traffic monitoring systems and improving system reliability and accuracy.

7. CONCLUSION

This paper proposed a vehicle behavior recognition algorithm for intelligent traffic monitoring systems, aimed at improving the accuracy and efficiency of vehicle behavior recognition. First, the overall framework of the algorithm was introduced, including two key parts: feature extraction and classification. Then, the paper detailed how to construct spatiotemporal feature bodies to better capture spatiotemporal characteristics in traffic monitoring videos. By analyzing the spatiotemporal relationships in video data, key features that represent vehicle behavior were extracted. In terms of classification, the paper used a multi-core SVM approach, leveraging the strong classification capability of multi-core SVM to accurately classify the extracted spatiotemporal features. Through a series of experiments, the proposed method demonstrated excellent performance across various datasets, including recognition accuracy for different algorithms with varying $q1$ weights, feature extraction results with different dimensions of fully connected layers, and results from different spatiotemporal feature extraction models.

Overall, the proposed intelligent traffic monitoring vehicle behavior recognition algorithm, through detailed spatiotemporal feature body construction and multi-core SVM classification, significantly improves vehicle behavior recognition accuracy. The experimental results validate the superior performance of the method across different datasets and complex traffic scenarios, demonstrating its potential application in intelligent traffic monitoring systems. This research has important value, first in its high accuracy, demonstrated by excellent recognition performance across multiple datasets, proving the method's effectiveness in capturing and recognizing vehicle behaviors in complex traffic environments. Secondly, the method shows strong adaptability, incorporating multiple feature extraction methods to handle a variety of traffic monitoring scenarios, exhibiting high generality.

However, there are some limitations. Due to the integration of various feature extraction and complex classification algorithms, the computational complexity is higher, which may impact real-time performance. Additionally, the performance of the algorithm depends on the diversity and quality of the training data, and differences in dataset distribution may affect the results. Future research can focus on optimizing the computational efficiency of the algorithm to enhance real-time processing capabilities, making it more practical for real-world traffic monitoring systems. Furthermore, integrating deep learning methods could further improve feature extraction and classification performance, particularly for handling large-scale and high-complexity traffic monitoring data. Exploring the integration of data from other sensors could also enhance the overall recognition capabilities of vehicle behavior and improve the robustness and accuracy of the system.

REFERENCES

[1] Liu, Y., Yang, C., Sun, Q. (2020). Thresholds based image extraction schemes in big data environment in intelligent traffic management. IEEE Transactions on Intelligent Transportation Systems, 22(7): 3952-3960. ttps://doi.org/10.1109/TITS.2020.2994386

[2] Lin, H.Y., Chang, C.C., Tran, V.L., Shi, J.H. (2020). Improved traffic sign recognition for in-car cameras. Journal of the Chinese Institute of Engineers, 43(3): 300-307. https://doi.org/10.1080/02533839.2019.1708801

[3] Li, J., Dong, Y. (2019). A new night traffic light recognition method. In Journal of Physics: Conference Series, IOP Publishing, 1176(4): 042008. https://doi.org/10.1088/1742-6596/1176/4/042008

[4] Xu, H., Srivastava, G. (2020). Automatic recognition algorithm of traffic signs based on convolution neural network. Multimedia Tools and Applications, 79(17): 11551-11565. https://doi.org/10.1007/s11042-019-08239-z

[5] Liu, B., Lam, C.T., Ng, B.K., Yuan, X., Im, S.K. (2024). A graph-based framework for traffic forecasting and congestion detection using online images from multiple cameras. IEEE Access, 12: 3756-3767. https://doi.org/10.1109/ACCESS.2023.3349034

[6] Wang, M., Liu, R., Yang, J., Lu, X., Yu, J., Ren, H. (2022). Traffic sign three-dimensional reconstruction based on point clouds and panoramic images. The Photogrammetric Record, 37(177): 87-110. https://doi.org/10.1111/phor.12398

[7] Dong, X., Lan, J., Wu, W. (2022). Research on real-time monitoring of video images of traffic vehicles and pedestrian flow using intelligent algorithms. International Journal of Advanced Computer Science and Applications, 13(12): 582-589. https://doi.org/10.14569/IJACSA.2022.0131271

[8] He, X., Li, L., Peng, H., Tong, F. (2024). A multi-level privacy-preserving scheme for extracting traffic images. Signal Processing, 220: 109445. https://doi.org/10.1016/j.sigpro.2024.109445

[9] Li, Y., Chu, L., Zhang, Y., Guo, C., Fu, Z., Gao, J. (2019). Intelligent transportation video tracking technology based on computer and image processing technology. Journal of Intelligent & Fuzzy Systems, 37(3): 3347-3356. https://doi.org/10.3233/JIFS-179137

[10] Amin, M.A., Hanif, M.K., Sarwar, M.U., Sarwar, M.K., Kanwal, A., Azeem, M. (2018). Video streaming analytics for traffic monitoring systems. International Journal of Advanced Computer Science and Applications, 9(11): 651-654.

[11] Mizutani, K. (2021). An extractive streaming system to reduce monitoring traffic. IEICE Communications Express, 10(1): 13-17. ttps://doi.org/10.1587/comex.2020XBL0121

[12] Chaudhary, S., Indu, S., Chaudhury, S. (2018). Video-based road traffic monitoring and prediction using dynamic Bayesian networks. IET Intelligent Transport Systems, 12(3): 169-176. https://doi.org/10.1049/iet-its.2016.0336

[13] Liu, G., Shi, H., Kiani, A., Khreishah, A., Lee, J., et al. (2021). Smart traffic monitoring system using computer vision and edge computing. IEEE Transactions on Intelligent Transportation Systems, 23(8): 12027-12038. https://doi.org/10.1109/TITS.2021.3109481

[14] Li, X.L. (2020). Big video monitoring scheme of traffic video based on Hadoop. Chinese Journal of Liquid Crystals and Displays, 35(11): 1204-1209.

[15] Pan, S., Li, P., Yi, C., Zeng, D., Liang, Y.C., Hu, G. (2020). Edge intelligence empowered urban traffic monitoring: A network tomography perspective. IEEE Transactions on Intelligent Transportation Systems, 22(4): 2198-2211. https://doi.org/10.1109/TITS.2020.3024824
[16] Al-qaness, M.A., Abbasi, A.A., Fan, H., Ibrahim, R.A., Alsamhi, S.H., Hawbani, A. (2021). An improved YOLO-based road traffic monitoring system. Computing, 103(2): 211-230. https://doi.org/10.1007/s00607-020-00869-8
[17] Barth, V., de Oliveira, R., do Nascimento, V. (2019). Vehicle speed monitoring using convolutional neural networks. IEEE Latin America Transactions, 17(6): 1000-1008. https://doi.org/10.1109/TLA.2019.8896823
[18] Li, D., Lasenby, J. (2022). A revised video vision transformer for traffic estimation with fleet trajectories. IEEE Sensors Journal, 22(17): 17103-17112. https://doi.org/10.1109/JSEN.2022.3193663
[19] Martín-Baos, J.Á., Rodriguez-Benitez, L., García-Ródenas, R., Liu, J. (2022). IoT based monitoring of air quality and traffic using regression analysis. Applied Soft Computing, 115: 108282. https://doi.org/10.1016/j.asoc.2021.108282