Vehicle Music Automation Control System Based on Machine Vision

Daqi Tian, Jinlin Chen, Xin Wang

***Abstract*—With the rapid development of Internet of Things technology, vehicle-based entertainment systems are developing towards a more natural, faster, and automated direction. To achieve automatic control of vehicle-based music, based on machine vision, the music automation control method is designed by recognizing human gestures. The improved gesture motion trajectory point localization method, and enhanced recognition method based on the HMM are proposed to control vehicle-based music. The research results indicated that the method proposed in this research had a recognition rate of 95.3% for non-similar gestures. The recognition rate for similar gestures reached 94.56%. This indicates that the designed method can accurately and automatically control in car music equipment It is expected that this research can provide some reference value for the current research on vehicle-based music automated control.**

***Index Terms*—Machine vision; Automation; HMM; Gesture recognition; Vehicle-based music system**

I. Introduction

SINCE the birth of computers, the interaction between humans and computers has always been centered around computers. Human beings actively learn computer language to achieve interpersonal interaction [1]. With the advancement of computer technology, the interpersonal interaction has developed to a stage based on graphical user interfaces. Computer control is achieved through image manipulation, but this interaction method is still not flexible enough [2-4]. The human-computer interaction in these two stages is mainly achieved through the mouse and keyboard, both of which belong to computer-centered contact interaction. With the further optimization of computer technology, the interaction between humans and computers has gradually shifted from computer-centered to human-centered. Interpersonal interaction is more natural and humane. Gesture recognition, facial recognition, speech recognition, etc. are already ubiquitous in daily life [5-6]. Based on the machine vision, Hidden Markov Model (HMM) is introduced to locate and recognize gesture trajectories, thereby achieving the vehicle-based music automated control. It is hoped that the research can provide some reference value for the development of vehicle-based music automation in China [7].

Manuscript received October 25, 2023; revised June 25, 2024.

Daqi Tian is a tutor of Henan Polytechnic Institute, Nanyang 473000, China. (corresponding author, e-mail: t1246185734@163.com)

Jinlin Chen is a tutor of Henan Polytechnic Institute, Nanyang 473000, China. (e-mail: a794722066@163.com)

Xin Wang is an engineer of State Gird Nanyang Electric Power Supply Company, Nanyang 473000, China. (e-mail: 736861042@qq.com)

II. Related Works

Gesture trajectory recognition is an important topic in human-computer interaction research. Accurate and natural gesture trajectory recognition can effectively improve the human-computer interaction efficiency. Gao P et al. proposed a video image motion recognition and multi-dimensional data capture model based on a deep learning framework. The Gaussian mixture model was used to extract the motion trajectory of the target. In deep learning features, the fusion of deep video features and video RGB tricolor features was considered a deep learning feature. Simulation experiments showed that this algorithm had high recognition accuracy for large-scale datasets and small gesture actions. The average classification accuracy was 85.79%. This algorithm ran at a speed of around 20 frames per second [8]. Gao Y Q et al. compared the processing method based on visual gesture recognition with the gesture recognition method based on wearable sensors. The results showed that visual gesture recognition methods had more natural and intuitive gesture trajectory recognition. Meanwhile, the challenges and future research directions of visual-based gesture recognition in tactile rendering were proposed [9]. Ding I et al. used the Leap Motion Controller image sensor to collect gesture data. Four different deep learning strategies for identity recognition based on gesture intention were proposed. Compared with traditional gesture recognition models, the proposed method had more prominent advantages [10]. Juan W combined gesture recognition with teacher classroom teaching. Then a dynamic gesture recognition method was proposed. The data in the gesture action area was converted into a grayscale image. Then the improved algorithm was used for data classification. The performance was analyzed through control experiments. The research results indicated that gesture recognition teaching in remote education could effectively improve educational efficiency, which also had higher accuracy [11]. Bose S R et al. utilized an optimized deep residual network model for real-time hand motion recognition. This method combined the RetinaNet model for hand detection with a deep separable convolution layer for precise gesture recognition. This model overcame the class imbalance problem encountered by traditional single stage hand motion recognition algorithms. The results showed that the accuracy and time consumption of this method were significantly improved [12].

The HMM used for images can effectively describe the statistical characteristics of multi-scale transformation coefficients and their correlations, which further utilizes the deep features of images. Wang X H et al. combined

multi-scale geometric analysis with HMM to discuss the effectiveness of the image processing. The results showed that this method could better describe the deep features of images [13]. Sebastien K et al. proposed a multi-level Bayesian HMM. The multivariate Poisson lognormal emission probability distribution, multi-level parameter estimation, and experimental specific conditional variables were combined. The results indicated that this multi-level Bayesian HMM framework was suitable for future research on the long-term plasticity of neural populations [14]. Aoudni Y et al. proposed an HMM for detecting and preventing attacks in cloud platforms. Firstly, the HMM was applied to detect attacks. Then, the k-means clustering-based deep learning model was used for attack recognition. The results showed that the model could effectively achieve attack detection on network platforms [15]. Paeng J W et al. proposed a new tracking algorithm based on factor HMM. In this framework, the geometric relationships between patches were encoded through interactive sampling or importance sampling on sets. The experimental results showed that this method outperformed other methods in both qualitative and quantitative aspects [16]. Huang M et al. proposed a new estimation method for non-parametric HMM. This estimation was based on a new composite likelihood method. Continuous observations were considered as independent random variables. Therefore, the model was transformed into a hybrid model. Then an improved Expectation Maximization algorithm was proposed to calculate the maximum composite likelihood. The results showed that this method had higher performance [17].

In summary, gesture trajectory recognition based on different algorithms has been widely researched. Simultaneously, the application fields are gradually expanding. The HMM has also been effectively applied in various data analysis. However, as the vehicle-based music continues to improve, higher implementation technologies are required for the automation control of vehicle-based music systems. Although research has been conducted on vehicle-based music systems, there is relatively little research on the application of gesture trajectory recognition systems combined with the HMM to the vehicle-based music automation control. Therefore, HMM is used to construct a gesture trajectory recognition system. Then the system is applied to the vehicle-based music system automation control. It is expected to build a simpler and more effective automatic control method for vehicle-based music.

III. Construction of Gesture Trajectory Localization and Recognition Method Based on Improved Attention Mechanism

A. Localization method design based on improved effective trajectory points

A very important issue in dynamic gesture recognition of vehicle-based music based on machine vision is effective trajectory point localization. Effective trajectory point positioning refers to the complete definition of a dynamic gesture from the beginning to the end. In machine vision sampling, it represents the hand motion trajectory from the beginning to the end of the gesture [18]. Leap Motion is a widely used gesture acquisition device currently. The device department determines the radius and center position information of the virtual sphere by analyzing the operator's palm and palm curvature, as shown in Fig. 1.

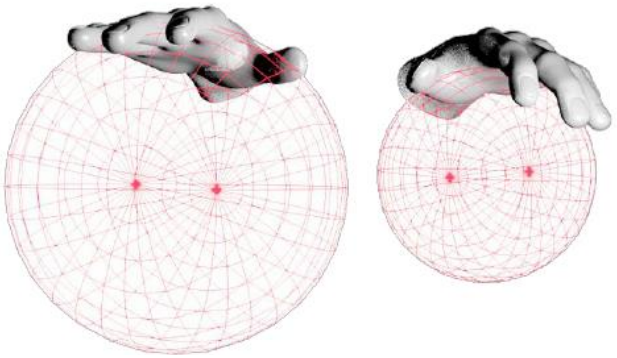


Fig. 1. Schematic diagram of Leap Motion detection virtual ball.

Fig. 1 is a schematic diagram of palm virtual ball detection. From this figure, as the curvature of the fingers gradually increases, the radius of the palm virtual ball gradually decreases until the radius decreases to zero during the palm closure. In this research, when the radius of the virtual ball in the palm is $R \leq 20$mm, the palm is in a closed state. When R>20mm, the palm is in an open state.

When performing dynamic gesture recognition, methods such as speed, distance threshold model, and motion region model are commonly used to locate the starting point of a trajectory [19]. However, these trajectory point positioning methods have certain limitations, because the operators are different. The gesture and movement habits of each operator are different. Therefore, the stability is not sufficient when collecting the operator's gesture trajectory. To solve the existing problems in the current trajectory point positioning method, a more specific finger state information than palm information is proposed. The main process is shown in Fig. 2.

Fig. 2 is the flowchart of an improved effective trajectory point localization algorithm. Compared with the existing trajectory point localization methods, the improvement is mainly reflected in determining the starting point of trajectory collection through finger state.

Everyone controls the opening or closing of their fingers in a similar way, eliminating personalized approaches that vary from person to person. Therefore, this improvement has higher robustness and stability compared with distance, speed, and region model methods. The movement speed of gestures at the starting point changes relatively slowly. Therefore, the sampling points at the initial position are prone to stacking and shaking, which directly affects the system recognition rate [20-21]. To eliminate the impact of external interference factors on recognition accuracy, the gesture trajectory data is pre-processed after collecting gesture trajectory data. The median filter is used to filter the motion trajectory of gestures. The median filter with a sliding window of 3×3 is used to filter the gesture trajectory data.

B Solid state feature extraction based on effective trajectory localization

Accurate feature extraction is the key to successful gesture recognition. The main task of the feature extraction

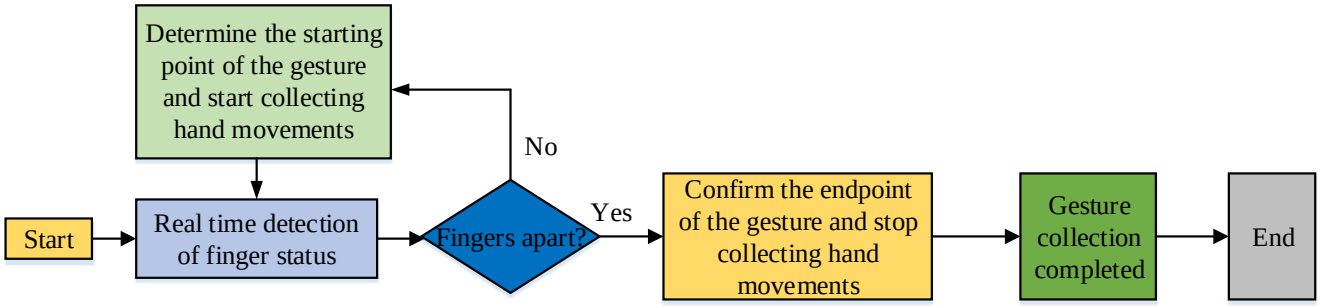
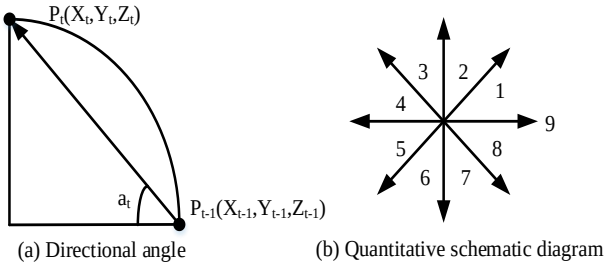


Fig. 2. Process of improved effective trajectory point localization algorithm.

stage is to extract essential features that can represent gesture trajectories from dynamic gestures, facilitating subsequent classification processing. When extracting dynamic features of gestures, the movement speed and amplitude of gestures vary among different individuals [22]. Therefore, the time and spatial differences in gesture trajectories should be eliminated. The parameters used to describe the motion trajectory of gestures include spatial position, motion rate, and motion direction angle. In the research, the directional angle information is used as a fixed feature for recognizing non-similar gestures.



(a) Directional angle (b) Quantitative schematic diagram

Fig. 3. Schematic diagram of directional angle and quantization on trajectory.

Fig. 3 shows the direction angle on the dynamic gesture motion trajectory and the quantification schematic diagram. After collecting gesture speed and coordinate information, adjacent point motion direction angle features are extracted. The directional angle is shown in Fig. 3 (a). The direction angle is expressed by the coordinate vectors of two adjacent moments and the positive X-axis counterclockwise. The coordinates of the two points are P_t and P_{t-1}. The azimuth angle a_t is expressed, as shown in formula (1).

$$\begin{cases} \arctan\left(\frac{z_t - z_{t-1}}{x_t - x_{t-1}}\right) * \left(\frac{180}{\pi}\right) + 180, \Delta < 0 \\ \arctan\left(\frac{z_t - z_{t-1}}{x_t - x_{t-1}}\right) * \left(\frac{180}{\pi}\right) + 360, \Delta z < 0 \\ \arctan\left(\frac{z_t - z_{t-1}}{x_t - x_{t-1}}\right) * \left(\frac{180}{\pi}\right), \Delta x > 0, \Delta z \geq 0 \end{cases} \quad (1)$$

To reduce computational complexity when quantifying the direction angle, the direction angle information obtained from formula (1) is quantified into 9 levels based on a 9-direction chain code. The quantified results are shown in Fig. 3 (b). In Fig. 3 (b), the directional information is evenly divided into 8 parts, represented by numbers 1-8. For example, when the directional angle is $a_t = 0$, it is encoded as 9. When $a_t \in \left(45^o, 90^o\right]$, the direction angle is encoded as 2. This digital encoding method represents feature vectors, which can facilitate subsequent recognition training.

C Gesture feature recognition model construction based on HMM

After extracting the gesture features, the final and most important step for gesture recognition is carried out. It is the gesture classification stage. The gesture classification method used in this research is the HMM, which is a doubly stochastic process. Each state is only related to the previous one. The output generated by this process is a sequence of states, which is independent of other previous states [23]. Another random process is used to describe the correspondence between states and observed values. The output generated by this process is a sequence of observed values. The basic schematic diagram of the HMM is shown in Fig. 4.

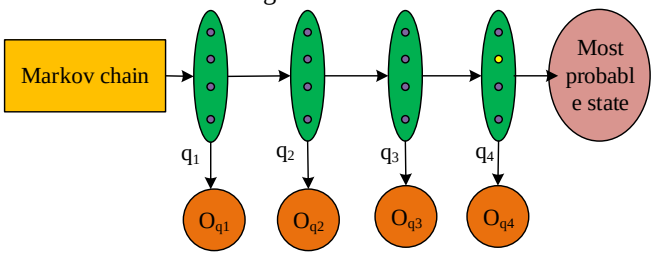


Fig. 4. Basic schematic diagram of HMM.

If S_i is the hidden state of the HMM topology model, the model contains a total of 6 hidden states. O_i represents the observed state of the HMM. In general, a model triplet can be used to describe the HMM, i.e. $\lambda = (A, B, \pi)$. A represents the probability matrix of state transition, which is the transition probability from one state to another. B represents the probability matrix of observed values, which is the distribution probability of observed state values in a certain state. π is the initial probability matrix, which is the probability in the initial state. The probability matrix A of state transition is shown in formula (2).

$$A = \begin{bmatrix} a_{ii} & 1-a_{ii} & & & & 0 \\ & a_{ii} & 1-a_{ii} & & & \\ & & \ddots & \ddots & & \\ & & & & a_{ii} & 1-a_{ii} \\ 0 & & & & & 1 \end{bmatrix} \quad (2)$$

In formula (2), the size of a_{ii} is related to the average duration d that each state can sustain. The relationship between the two is shown in formula (3).

$$a_{ii} = 1 - \frac{1}{d}, d = \frac{\bar{T}}{N} \quad (3)$$

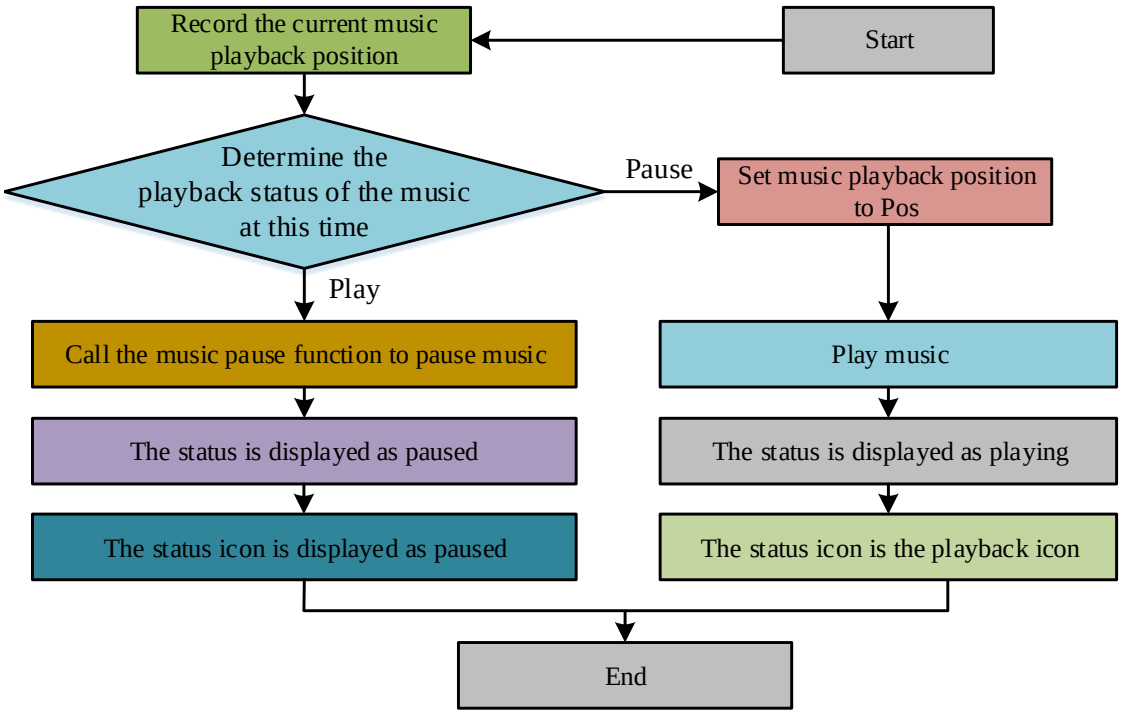


Fig. 5. Implementation process of music playback and pause.

In formula (3), $\bar{T}$ represents the average sample length, which is the average length of all training samples in the corresponding gesture. If the probability of each state occurring is the same, then formula (3) can be used to represent the observation probability matrix B, as shown in formula (4).

$$B=\{b_{jk}\}, b_{jk}=\frac{1}{M} \tag{4}$$

In formula (4), $j=1,2,\ldots,N; k=1,2,\ldots M$. The feature vectors in this research are obtained through 9-direction chain code quantization. Therefore, $M=16$. The HMM topology structure used in this research is a left and right side-band model. Therefore, the initial probability matrix π is assigned by formula (5).

$$\pi=[1\ 0\ \cdots\ 0]^T \tag{5}$$

After determining the initial model of the algorithm, the Baum-Welch algorithm is used to train the HMM. During the training process, the new model parameter $\bar{\lambda}=(\bar{A},\bar{B},\bar{\pi})$ is continuously optimized. If the output probability $|P(O|\bar{\lambda})-P(O|\lambda)|\le 1\times 10^{-4}$ is met, the output probability converges and the iterative operation can be stopped.

When performing gesture recognition, it is inevitable to encounter familiar gestures, such as gestures "0" and "O". The HMM classification algorithm used in the above research classifies dissimilar gestures without optimizing for similar gestures, resulting in a low recognition rate for similar gestures. On the basis of extracting fixed features in the above chapters, the secondary feature extraction is carried out. The recognition effect is enhanced based on a two-layer recognition system.

The percentage of sub-regions, the aspect ratio of trajectory regions, and the number of corners are all secondary feature contents. The corner points can also be called extreme points. It can represent points located on different things in two adjacent directions, points with special attributes, or points with artificial definitions. In the gesture trajectory, the motion direction angle difference value Δa_t between adjacent two points is expressed as formula (6).

$$\Delta a_t > T, \Delta a_t = a_t - a_{t-1} \tag{6}$$

In formula (6), T represents the angle threshold. When Δa_t is greater than the corresponding gesture angle threshold, the current point is a corner point. The percentage of sub-regions is required to be solved. Firstly, the gesture trajectory region is evenly divided into multiple sub-regions. Then the ratio of sub-regions to the total region is calculated, as shown in formula (7).

$$subArea_i=\frac{point\_num_i}{point\_num}\times 100\%, i=1,2,3,4 \tag{7}$$

In formula (7), $point\_num_i$ represents the number of sampling points in the i -th sub-region. $point\_num$ represents the number of points used in the gesture trajectory. Each gesture can obtain a vector group of sub-region percentages. According to formula (8), the corresponding sub-region percentage standard value of the gesture is obtained.

$$subArea_i^*=\frac{1}{k}\sum_{i=1,j=1}^{i=4,j=k} subArea_{i,j} \tag{8}$$

The percentage of corner sequence numbers refers to the percentage of the corner sequence corresponding to a certain gesture compared to the total number of sampling points in the entire gesture trajectory. If there is more than one corner in a gesture trajectory, the percentage of each corner in the graph needs to be calculated. Finally, the average value obtained is used to represent the standard value of the corner number percentage. The ratio of length to width in the trajectory area is one of the secondary features.

Formula (9) is used to represent the aspect ratio of the trajectory area.

$$Area\_tatio = \frac{H}{W} \tag{9}$$

The secondary feature extraction can enhance the detailed information of gesture features. In this enhanced recognition process, if the input gesture action still belongs to a similar gesture set after initial recognition by the HMM, the current input gesture is continued to undergo secondary feature extraction to obtain enhanced recognition. If the HMM recognition result belongs to a set of similar gestures, the corner points of the current input gesture are extracted. The gesture set with corner point features similar to the output gesture is matched. If the matching result based on the number of corner points does not meet the requirements, the sub-region percentage feature is calculated until the model outputs a unique value. Finally, the algorithm ends.

To ensure the performance of the vehicle-based music control system, i.MX6 Quad is used to develop vehicle-based music systems. This embedded development speed is fast and the functions are guaranteed [24]. After developing the vehicle-based music system, corresponding software and plugins are installed in the system. Taking the music playback and pause of the vehicle-based music system as an example, the implementation process is shown in Fig. 5.

In the above process, the Qt WebSockets is used to design a vehicle-based music player to achieve signal connection between gesture actions and playback slots. The conversion between gesture actions and the vehicle-based music module is successfully completed.

IV Performance Analysis of Gesture Trajectory Localization and Recognition Methods Based on Machine Vision

A. Experimental analysis of effective trajectory point localization

To verify whether the proposed gesture recognition method based on machine vision is suitable for automotive music automation control, corresponding experiments are conducted to analyze it. To fully validate the effectiveness of the proposed method, a total of 36 gestures are used, represented by the numbers "0-9" and "A-Z", respectively. Each gesture corresponds to different functions in vehicle-based music playback. For example, "1" corresponds to "next one" in the music playback system. "S" represents "stop playing music" in the music playback system. The other numbers and letters correspond to one function, respectively. The three-dimensional motion trajectory corresponding to each gesture is shown in Fig. 6. The yellow dots in the figure represent the starting point of the gesture motion trajectory. The red dot represents the endpoint of the gesture motion trajectory.

Comparative experiments are conducted between the unimproved trajectory localization method and the improved effective trajectory point localization proposed in this research, including distance model and speed threshold model. Three methods are evaluated based on the trajectory success rates. In the experiment, three volunteers randomly select 10 gestures from 36. The trajectory point positioning experiment is carried out. Fig. 7 shows the effective trajectory point localization experiment results. The experimental results show the success rates of 10 gestures for volunteers A, B, and C. Fig. 7 (a), (b), and (c) show the success rate results for different methods. Among them, Fig. 7 (a) shows the success rate of effective trajectory point localization based on the speed threshold model. From the experimental results, the positioning success rate varied greatly, which was not stable. The maximum difference between the highest and lowest success rates was 31%, with an average success rate of 58%. Fig. 7 (b) shows the success rate of effective trajectory point localization based on the distance threshold model. From the experimental results, the highest localization success rate of this method was 82%. The minimum success rate was 49%, and the average success rate was 69.7%. Fig. 7 (c) shows the success rate of effective trajectory point localization based on the method proposed in this study. From the experimental results, the positioning success rate of this method was relatively stable, without significant fluctuations. The interpolation between the highest and lowest success rates was only 9%, with an average success rate of 96%. Fig. 7 (d) shows the average success rate of the three methods. From this comparison chart, the average success rate of the proposed method was the highest. Compared with the other two methods, it had higher stability, lowest dispersion, and best robustness.

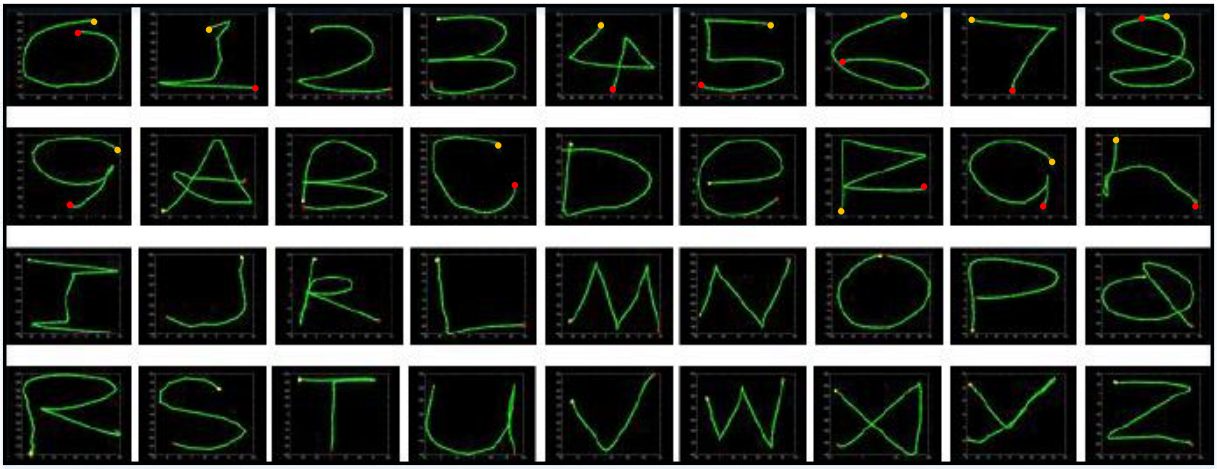


Fig. 6. 3D motion trajectories corresponding to each gesture.

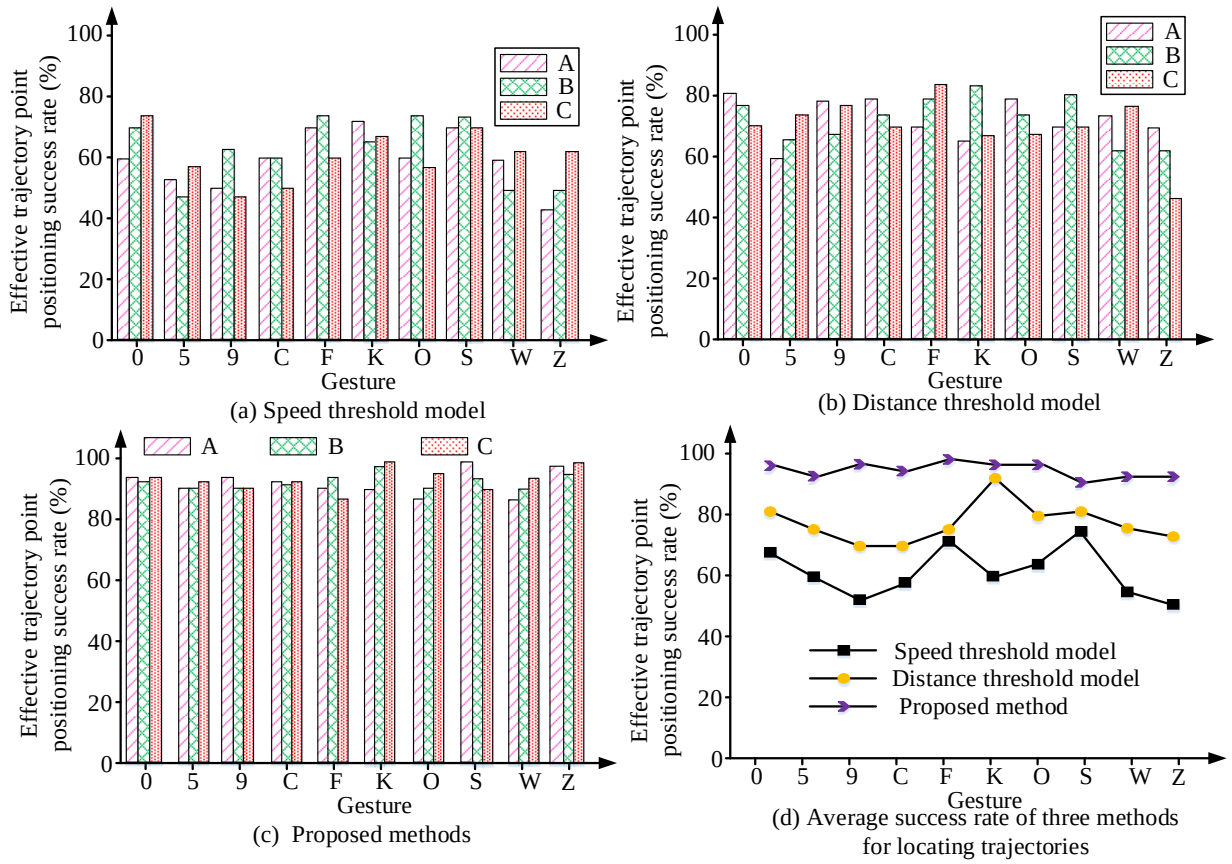


Fig. 7. Experimental results of trajectory positioning.

B Recognition results analysis based on HMM

The performance of the HMM proposed in the research is verified. Then, it is validated using different datasets, including 36 gesture actions and the American Sign Language (ASL) gesture image dataset used in the aforementioned study. This dataset contains a total of 120000 images. The proposed HMM-based gesture trajectory recognition method (Algorithm 1), the gesture recognition method based on convolutional neural network (Algorithm 2), and the gesture recognition method based on Kinect depth image (Algorithm 3) are compared. The Mean Intersection over Union (mIoU) represents the ratio of the intersection and union of the calculated and actual results of the model, which can be used as an evaluation standard for gesture segmentation accuracy. The mIoU of the three models is shown in Fig. 8. In Fig. 8 (a), the differences in mIoU values among the three models were significant. The Algorithm 1 had the smallest fluctuation in mIoU values, with multiple maximum and minimum values appearing throughout the process. The final mIoU value was 0.83. The mIoU value of Algorithm 2 after convergence was 0.79. The mIoU of Algorithm 3 was 0.75. The proposed method had a higher mIoU value of 0.04 and 0.08 compared with Algorithm 2 and Algorithm 3, respectively. In Fig. 8 (b), the mIoU values of the three methods were 85.6%, 84.3%, and 82.1%, respectively. In the ASL dataset, the performance of Algorithm 1 was also higher than its comparison methods. This indicates that the proposed method has better recognition performance for gesture actions.

The convergence effects of the three methods during operation are compared. The loss values are shown in Fig. 9. The converged loss value of the proposed method was 1.6, while the converged loss values of Algorithm 2 and Algorithm 3 were 2.4 and 2, respectively. The method proposed in the study has lower loss values, which has better convergence performance.

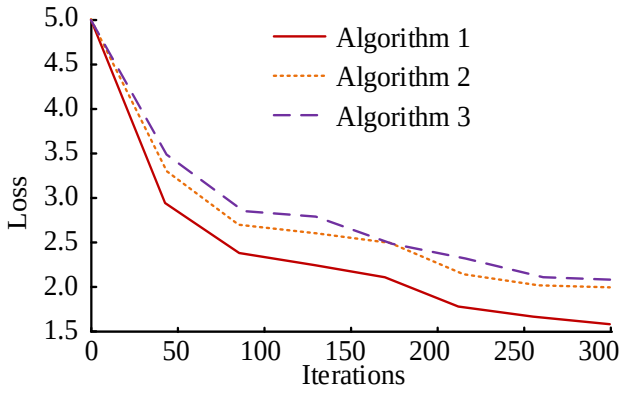


Fig. 9. Curve of loss value with number of iterations.

The precision, Recall and F1-value are varied. The results are shown in Fig. 10. In Fig. 10 (a), the accuracy values of the three methods were 96.12%, 94.33%, and 92.87%, respectively. In Fig. 10 (b), the recall values of the three methods were 97.56%, 95.43%, and 92.67%, respectively. In Fig. 10 (c), the F-values of the three methods were 96.86%, 93.59%, and 91.35%, respectively. The performance of the HMM-based gesture trajectory recognition method used in the research is significantly superior to other commonly used methods.

To verify the recognition effect of gesture trajectories proposed in the research, the accuracy of different gesture samples is analyzed. The results are shown in Table I. The

gesture samples involved in the test include up, down, left, right, and fist gestures. From Table I, in the training and testing samples, the fist sample had the highest recognition accuracy, reaching 91% and 96.3%, respectively. This is because the action characteristics are more pronounced. The up and down actions only differed in direction. The commonality between the two is higher. Therefore, the accuracy is slightly lower.

To verify the effectiveness of the designed method, confusion matrix experiments are conducted on the dataset. The confusion matrix results are shown in Fig. 11. From Fig. 11, the recognition accuracy of up, down, left, right, and fist were 92%, 94%, 95%, 91%, and 91%, respectively. The recognition accuracy of all actions was above 90%. The above indicates that the algorithm proposed in the study has good action recognition performance.

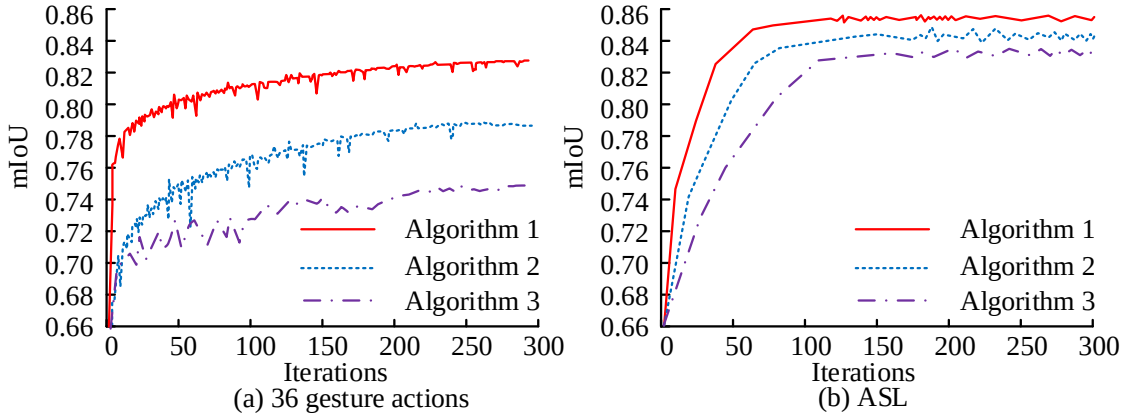


Fig. 8. Comparison of mIoU curves for three models.

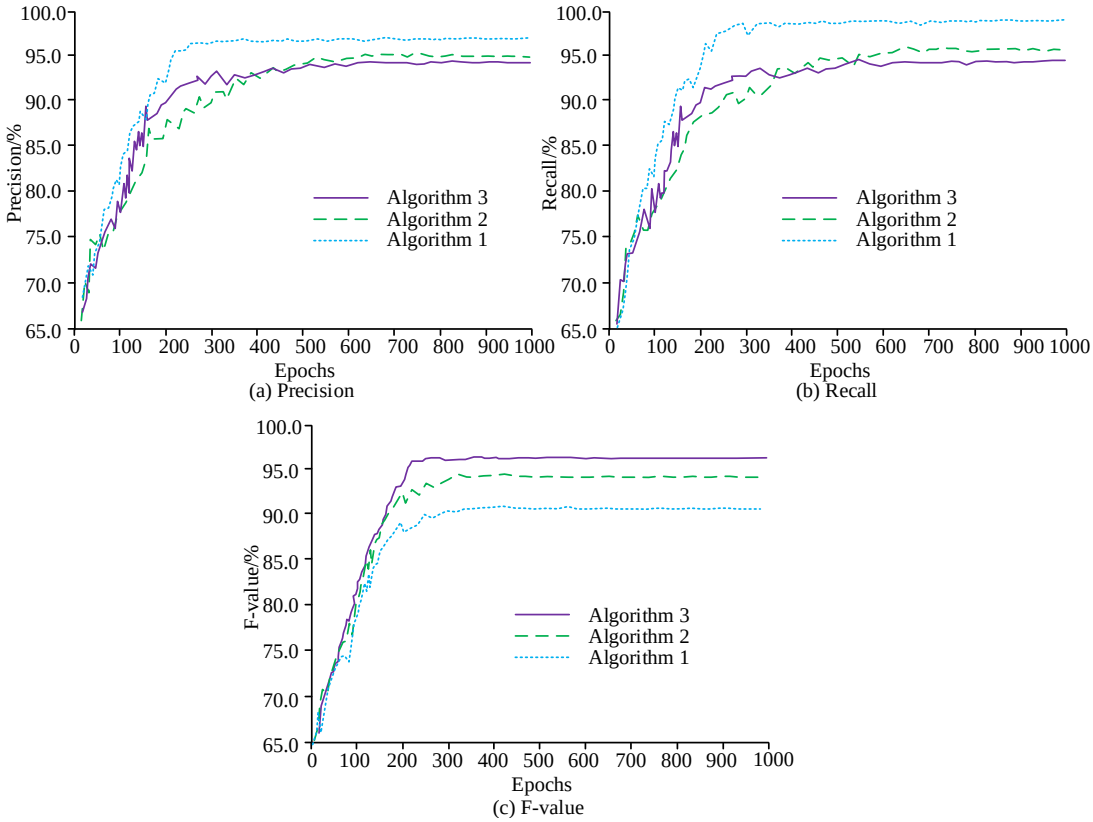


Fig. 10. Performance Comparison of Different Methods.

TABLE I
RECOGNITION RATES OF DIFFERENT GESTURE TYPES

Gesture type	Training sample size	Correct recognition quantity	Recognition rate	Testing sample size	Correct recognition quantity	Recognition rate
Up	100	85	85%	300	267	89%
Down	100	71	71%	300	273	91%
Left	100	89	89%	300	264	88%
Right	100	82	82%	300	287	95.7%
Fist	100	91	91%	300	289	96.3%

Fig. 11 Confusion Matrix Experimental Results.

Experiments are conducted on the recognition success rate of HMM. If a gesture is recognized as another gesture more than three times in an experiment, then these two gestures are considered similar. The initial recognition results based on HMM obtained in this research are shown in Fig. 12. The horizontal axis in the figure represents the various gestures in the research. The vertical axis represents the recognition accuracy. From Fig. 12, the recognition method based on HMM had higher recognition rate for non-similar gestures, with an average accuracy of 98.2%. It shows an excellent stability. The recognition method based on HMM had lower success rate in recognizing similar gestures. The average value was only 67.5, demonstrating the poor stability.

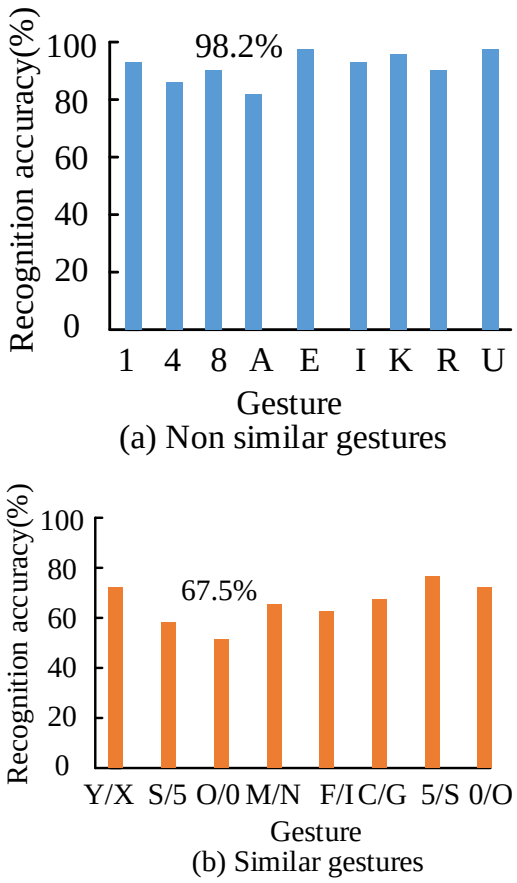


Fig. 12. Initial recognition results based on HMM.

The recognition of similar gestures is further enhanced on the basis of HMM. Fig. 13 shows the comparison between the initial recognition and enhanced recognition of similar gestures. From the above, the initial recognition rate based on HMM was relatively low. After enhanced recognition, the accuracy of gesture recognition was significantly improved, with an average recognition rate of 18%. The final recognition rate of the system for similar gestures was as high as 94.56%. In summary, the automatic control method based on machine vision for vehicle-based music proposed in this research can achieve various control operations for music status through gestures.

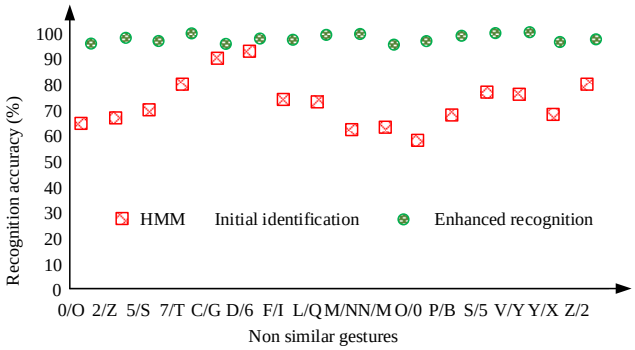


Fig. 13. Comparison of initial recognition and enhanced recognition for similar Gestures.

C Application effect of vehicle music automation system based on gesture recognition

The physiological data of various gestures performed by users are analyzed. The electromyographic data of the tester is used as an example for discussion. According to the sample amplitude of the gesture trajectory for each tester operating the vehicle-based music activation function, Delsys electromyography testing software is used for testing. After analyzing the data, the trajectory samples of the final 20 testers are obtained when operating 15 vehicle-based music activation functions. The results are shown in Table II. The sampling operation of the vehicle-based music activation function trajectory is repeated. Each different gesture trajectory sample presents different electromyographic data. The generated EMG values were relatively high. The results show that various gesture movements for vehicle-based music automatic control do not cause excessive fatigue and muscle stiffness for users.

TABLE II
AVERAGE EMG VALUES OF THE VEHICLE-BASED MUSIC ACTIVATION FUNCTION TRAJECTORY SAMPLE

Sample 1	0.0000086254
Sample 2	0.0000098643
Sample 3	0.0000074659
Sample 4	0.0000085626
Sample 5	0.0000098553
Sample 6	0.0000095406
Sample 7	0.0000089673
Sample 8	0.0000092137
Sample 9	0.0000088631
Sample 10	0.0000095824
Sample 11	0.0000077961
Sample 12	0.0000071076
Sample 13	0.0000094628
Sample 14	0.0000097142
Sample 15	0.0000089807

V Conclusion

The human-centered interpersonal interaction model is the future development trend. On the basis of machine vision, the vehicle-based music automatic control was achieved through human gesture recognition. Based on the improved traditional effective trajectory point localization method, the localization success rate was improved. On the basis of HMM, enhanced recognition was adopted to improve the recognition rate of similar gestures, thereby more accurately controlling the music system state. Subsequently, the proposed method was validated through experiments. The research results indicated that the improved effective trajectory point positioning method had an accuracy of 95.3%. The accuracy of non-similar gesture recognition was 98.2%. The accuracy of similar gesture recognition was 94.56%. The analysis results of different datasets showed that the proposed method also had higher recognition accuracy. In the confusion matrix experiment, the recognition accuracy of up, down, left, right, and first were 92%, 94%, 95%, 91%, and 91%, respectively. The above indicates that the algorithm proposed in the study has good action recognition performance. This proves the practicality of the proposed method. However, there are still some shortcomings in this research. The topology structure of the HMM used in the research can be further studied. It is hoped that improvements can be made in future research.

References

[1] S. Saqib, A. Ditta, M. A. Khan, S. A. R. Kazmi, "Intelligent dynamic gesture recognition using CNN empowered by edit distance," *CMC-computers Materials & Continua*, vol. 66, no. 2, pp. 2061-2076, 2020.

[2] Y. L. Chung, H. Y. Chung, W. F. Tsai, "Hand gesture recognition via image processing techniques and deep CNN," *Journal of Intelligent and Fuzzy Systems*, vol. 66, no. 2, pp. 1-14, 2020.

[3] S. Saee, A. Abdullah, N. Z. Jhanjhi, M. Naqvi, M. Masud, M. Alzain, "Hybrid grabcut hidden Markov model for segmentation," *CMC-computers Materials & Continua*, vol. 7, no. 2, pp. 851-869, 2020.

[4] J. Murphy, "Hand gesture system enhances human-computer interaction," *Laser Focus World: The Magazine for the Photonics & Optoelectronics Industry*, vol. 58, no. 2, pp. 14-15, 2022.

[5] N, Yadav, A. K. Singh, S. Pal, "Improved self-attentive musical instrument digital interface content-based music recommendation system," *Computational Intelligence*, vol, 38, no. 4, pp. 1231-1257, 2022.

[6] Fan Yue, Shijian Huang, Qingming Chen, Siyan Hu, Yong Tan, Suihu Dang, and Derong Du, "A Novel Two-stream Architecture Fusing Static and Dynamic Features for Human Action Recognition," IAENG International Journal of Computer Science, vol. 50, no.2, pp394-401, 2023

[7] Andi W. R. Emanuel, Paulus Mudjihartono, and Joanna A. M. Nugraha, "Snapshot-Based Human Action Recognition using OpenPose and Deep Learning," IAENG International Journal of Computer Science, vol. 48, no.4, pp862-867, 2021

[8] P. Gao, D, Zhao, X. Chen, "Multi-dimensional data modelling of video image action recognition and motion capture in deep learning framework," *IET Image Processing*, vol. 14, no. 7, pp. 1257-1264, 2020.

[9] Y. Q. Gao, X. Lu, J. B. Sun, X. L. Tao, X. M. Huang, Y. X. Yan, J. Liu, "Vision-based hand gesture recognition for human-computer interaction-A Survey," *Wuhan University Journal of Natural Sciences*, vol. 25, no. 2, pp.169-184, 2020.

[10] I. Ding, N. W. Zheng, M. C Hsieh, "Hand gesture intention-based identity recognition using various recognition strategies incorporated with VGG convolution neural network-extracted deep learning features," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 4, pp. 7775-7788, 2021.

[11] W. Juan, "Gesture recognition and information recommendation based on machine learning and virtual reality in distance education," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 4, pp. 7509-7519, 2021.

[12] S. R. Bose, V. S. Kumar, "In-situ identification and recognition of multi-hand gestures using optimized deep residual network," *Journal of Intelligent & Fuzzy Systems*, vol. 41, no. 6, pp. 6983-6997, 2021.

[13] X H. Wang, R. X. Song, S. J. Qu, Z. H. Mu, C. M. Song, "Advance in multiscale geometric analysis image hidden Markov tree model," *Acta Electronica Sinica*, vol. 50, no. 1, pp. 238-249, 2022.

[14] K. Sebastien, M. M. Sebastian, C. Gino, B. Marco, F. Pier, A. Emmeke, B. James, "Bayesian multilevel hidden Markov models identify stable state dynamics in longitudinal recordings from macaque primary motor cortex," *The European Journal of Neuroscience*, vol. 58, no. 3, pp. 2787-2806, 2023.

[15] Y. Aoudni, C. Donald, A. Farouk, K. B Sahay, D. V. Babu, V. Tripathi, D. Dhabliya, "Cloud security based attack detection using transductive learning integrated with hidden Markov model," *Pattern Recognition Letters*, vol. 157, no. 5, pp. 16-26, 2022.

[16] J. W. Paeng, J. Kwon, "Visual tracking using interactive factorial hidden Markov models," *IET Signal Processing*, vol. 15, no. 6, pp. 365-374, 2021.

[17] M. Huang, Y. Huang, W. Yao, "Statistical inference for the nonparametric and semiparametric hidden Markov model via the composite likelihood approach," *Science China Mathematics*, vol. 66, no. 3, pp. 601-626, 2023.

[18] A. Taguchi, K. Hara, R. Ikesu, K. Kawana, T. Tsuruga, J. Tomio, Y. Osuga, "Prognosis of high-risk human papillomavirus-related cervical lesions: A hidden Markov model analysis of a single-center cohort in Japan," *Cancer Medicine*, vol. 11, no. 3, pp. 664-675, 2022.

[19] F. Shehzad, M. Khan M, M. Asfand, M. Sharif, M Alhaison, U Tariq, A. Majumdar, O. Thinnukool, "Two-stream deep learning architecture-based human action recognition," *Computers, Materials Continua*, vol. 74, no. 3, pp. 5931-5949, 2022.

[20] A. T. Salawudeen, P. J. Nyabvo, H. U. Suleiman, I. S. Momoh, E. K. Akut, "Heuristic hidden Markov model for fuzzy time series forecasting," *International Journal of Intelligent Systems Technologies and Applications*, vol. 20, no. 2, pp. 146-166, 2021.

[21] F. Bartolucci, A. Farcomeni, "A hidden Markov space-time model for mapping the dynamics of global access to food," *Journal of the Royal Statistical Society: Series A Statistics in Society*, vol. 185, no. 1, pp. 246-266, 2022.

[22] S. Li, J. Lian, "Hidden Markov model based control for singular Markov jump systems," *Journal of the Franklin Institute*, vol. 358, no. 3, pp. 4141-4158, 2021.

[23] B. Matuszewski, "A web-based framework for distributed music system research and creation," *Audio Engineering Society*, vol. 68, no. 10, pp. 717-726, 2020.

[24] A. K. Harshin, "Interface adapted LBB-stable finite elements on fluid structure interaction problems in fully Eulerian framework," *Applied Numerical Mathematics*, vol. 162, no. 2021, pp. 283-300, 2021.