Research Article

Zhehua Fan and Fengqun Huang*

Tai Chi movement segmentation and recognition on the grounds of multi-sensor data fusion and the DBSCAN algorithm

https://doi.org/10.1515/nleng-2025-0151
received September 9, 2024; accepted May 16, 2025

Abstract: Tai Chi is a traditional Chinese martial art with unique movements and philosophical connotations. The research on motion segmentation and recognition in Tai Chi is of great significance for the learning and teaching of Tai Chi. To address such issues, this study introduced multi-sensor data fusion (MSDF) methods and density-based noise application spatial clustering algorithms. First, the principle and process of multi-sensor stereo matching and data fusion were introduced. Then, the principle of minimum description length was adopted to capture feature points of its trajectory and achieve trajectory cutting. Finally, clustering algorithms were used for clustering to further achieve accurate action recognition. The results showed that in the ten movements recognition experiments of Wu-style Tai Chi, under binary classification, the average action recognition accuracy of the proposed algorithm reached 98.52%, which was 5.43 and 2.94% higher than other advanced algorithms, respectively. In the scenario of multi-classification, the average accuracy of the proposed algorithm was as high as 92.33%, which was significantly better than other algorithms. This indicated that the Tai Chi action segmentation and recognition method on the grounds of the fusion of the MSDF method and a density-based noise application spatial clustering algorithm had significant performance advantages. It provided more accurate and real-time technical support for intelligent teaching, training, and competition of Tai Chi.

Keywords: multiple sensors, data fusion, DBSCAN algorithm, Tai Chi movements, segmentation and recognition

*** Corresponding author: Fengqun Huang,** Ministry of Sports, Xiamen Institute of Technology, Xiamen, 361021, China, e-mail: huangfengqun8866@163.com
Zhehua Fan: Physical Education College, Putian University, Putian, 351100, China

1 Introduction

With the increasing awareness of health, Tai Chi, as a traditional fitness exercise, has received widespread attention. To accurately evaluate and guide Tai Chi practice, motion segmentation and recognition technology have become a research hotspot. The traditional methods for recognizing Tai Chi movements mainly rely on video analysis or manual observation, which have limitations in accuracy and real-time performance. Video analysis is susceptible to interference from factors such as lighting changes, background noise, camera angles, and target occlusion, which can affect the accuracy of recognition [1]. However, manual observation methods are influenced by subjective factors, and there may be significant differences in the evaluation results between different observers, making it difficult to ensure the objectivity and consistency of the evaluation. In addition, traditional action recognition techniques usually only focus on a single type of sensor data, such as accelerometers or gyroscopes, and ignore the important role of multi-sensor information fusion in action recognition [2,3]. On the grounds of this type of problem, we introduced multi-sensor data fusion (MSDF) and density-based spatial clustering of applications with noise (DBSCAN) to capture multi-dimensional action information of practitioners and improve the accuracy of action recognition. Among them, MSDF achieves multi-dimensional capture of motion information by integrating visual and inertial sensor (ISE) data. Visual sensors accurately capture the spatial and shape features of actions based on video data. ISEs detect real-time changes in body acceleration, rotation angle, and direction, comprehensively tracking the motion status of various parts of the body. The DBSCAN algorithm can automatically identify cluster structures based on data density, and noisy data have better robustness.

The novelty of this research is reflected in the following four key aspects: first, MSDF and DBSCAN algorithms were deeply integrated and applied to the field of

Tai Chi movement segmentation and recognition, opening up a new path for research in this field. Second, a triangular layout multivision sensor sampling system consisting of three Kinect cameras was constructed, which utilizes multithreading technology to achieve synchronous sampling and timestamp correlation. Third, the DBSCAN algorithm was improved by proposing an IST-DBSCAN algorithm based on spatiotemporal distance, combining sample statistical characteristics with contour coefficient optimization parameters, and introducing particle filters to effectively enhance the algorithm's ability to process nonlinear trajectory data in complex environments. Fourth, a user-friendly interactive system was developed that provided real-time visual feedback, personalized progress tracking, and multimodal feedback functions, achieving a close integration of technical research and practical applications.

The research includes four sections. Section 2 provides an overview of motion recognition and sensor fusion in Tai Chi. Section 3 introduces the principles and processes of MSDF and DBSCAN algorithms. Section 3.1 discusses the stereo matching and data fusion of MSDF, and Section 3.2 introduces the DBSCAN clustering algorithm to achieve the segmentation and recognition of Tai Chi movement trajectories. Section 4 conducted experimental verification on the proposed method in the research. Section 5 summarizes and discusses the experiment, and provides prospects for future research.

2 Related works

At present, the motion segmentation and recognition technology of Tai Chi is a hot topic in sports. Xu *et al.* and other scholars proposed a supervised Tai Chi action sequence segmentation method on the grounds of trajectory primitives and geometric features for the segmentation problem of complex action sequences in human activities. By learning trajectory primitives through unsupervised clustering and extracting geometric features, relevant outcomes show that this method achieves state-of-the-art performance [4]. Aiming at solving the problem of low accuracy in recognizing dynamic action sequences, researchers such as Mulcahy *et al.* proposed a method of using deep learning algorithms to recognize Tai Chi movements. It uses deep learning models to extract motion features from the spatiotemporal trajectories of human joints to recognize Tai Chi movements. The outcomes showcase that the gesture recognition rate of this deep learning algorithm reaches 89.22%, and the false detection rate is significantly reduced [5]. Regarding the issue of coordination and variability between lower limb joints in elderly Tai Chi exercises, Yan *et al.* and other researchers used the Vicon 3D motion capture system and recruited 30 female Tai Chi practitioners for the experiment. The results show that in Tai Chi exercise, the continuous relative phase changes of the hip, knee, and ankle segments are frequent, and the coordination amplitude and coordination variability between joints are lower than those of normal walking [6]. Liu *et al.* proposed an action segmentation algorithm on the grounds of encoding and decoding, and global temporal information to address the issues of prediction errors and decreased segmentation quality caused by over-segmentation in existing action segmentation algorithms. The proposed algorithm utilizes a long short-term memory neural network for capturing global timing information. The results showed that the algorithm achieved a frame accuracy of 93% on the constructed real Tai Chi action dataset [7].

Multi-sensor fusion is extensively utilized in many aspects of society. Chai *et al.* and other scholars proposed a method on the grounds of multi-sensor fusion for indoor mapping and positioning, which constructs a sparse single lane semantic map by fusing waypoints, semantic landmarks, and Wi Fi landmarks. The results indicate that this method has high map quality and positioning accuracy in different scenarios [8]. Researchers such as Sengupta *et al.* found a lack of a large number of labeled datasets in supervised learning, and therefore proposed a high-precision camera object detection and joint calibration method on the grounds of YOLOv3. The results indicate that this method is efficient and easy to implement, and can rapidly develop multi-sensor datasets [9]. Li *et al.* proposed an obstacle detection and tracking method on the grounds of multiple LiDARs to address issues such as difficulty distinguishing adjacent obstacles and difficulty tracking occluded obstacles. This method uses an adaptive voxel grid DBSCAN algorithm and a region growing algorithm to detect obstacles. The outcomes showcase that the average detection accuracy reaches 97.53%, the average tracking accuracy is 95.1%, and the entire process only takes 30 ms [10]. Purohit *et al.* have proposed a modular real-time multi-sensor fusion framework to address the issue of environmental perception in intelligent transportation applications. By collaborating with multiple sensors, this framework can improve the accuracy and probability of perception. The results show that the calculation time of the framework is less than 10 ms, and it can detect changes in external calibration of sensors and potential sensor failures [11]. To solve the problem of wheat ear localization in dense crop scenes, Zhang *et al.* proposed a three-dimensional object detection method on the grounds of sensor fusion. This method combines data from binocular cameras and LiDAR to generate two-dimensional bounding boxes through visual detection. The results show that this method can accurately and robustly detect wheat spikes in complex environments and obtain high-precision wheat density information [12].

In summary, researchers both domestically and internationally have conducted extensive research on motion recognition in Tai Chi and have achieved certain results. To further improve the performance of action recognition, this study introduces a combination of MSDF and DBSCAN algorithms, which are jointly applied to the action segmentation and recognition of Tai Chi. This is to provide reference and inspiration for the action analysis of other traditional sports projects and promote the intelligent development of the sports field.

3 Action recognition on the grounds of MSDF and DBSCAN algorithm

On the grounds of the characteristics of Tai Chi movement, this section first introduces the principle and process of multi-sensor stereo matching and data fusion, where multi-sensors include visual and ISE. Subsequently, in response to the problem of Tai Chi action recognition, the minimum description length (MDL) principle was introduced to capture feature points of its trajectory and achieve trajectory segmentation. Finally, the DBSCAN algorithm is used for clustering to further achieve accurate action recognition.

3.1 Vision ISE stereo matching and data fusion

In Tai Chi exercise, due to the presence of a large number of turning movements, relying solely on a single visual sensor for data collection may not be able to obtain complete data in certain situations, such as when the subject is facing away from the sensor or joint points are obstructing each other. To address this issue, the study constructed a multivision sensor sampling system consisting of three Kinects, ensuring that all joint points can be captured by at least one camera at any time. The study utilizes multi-threaded tools to achieve simultaneous sampling of three cameras and real-time recording of the correspondence between each sampling point and system time [13,14].

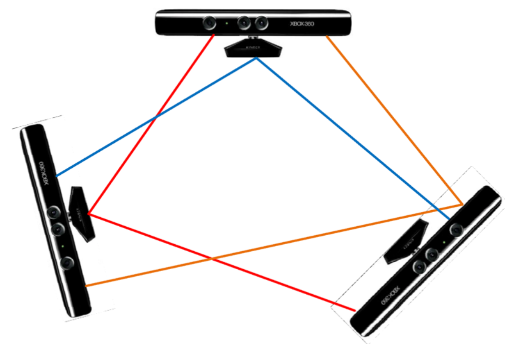

Figure 1: Schematic diagram of the three Kinect sampling.

The relevant structure of the three Kinect sampling is shown in Figure 1.

In Figure 1, three Kinect sensors form a triangular layout, and no matter how the practitioner moves or turns, at least one Kinect can capture their movements. The data from each Kinect are connected to the central processing unit through a cable, and synchronized sampling is achieved using multi-threaded tools to ensure data consistency and accuracy. Each sampling point is associated with the system time, facilitating time alignment in subsequent processing. Traditional action recognition research often relies on a single visual sensor, which is susceptible to factors such as viewing angle and occlusion during the acquisition process, resulting in data loss. The triangular layout sampling system constructed in this study, consisting of three Kinect cameras, can ensure that at any time, all the joints of the practitioner can be captured by at least one camera, effectively solving the problem of data collection for Tai Chi turning movements. Due to the differences in the coordinates of the joint points collected by ISEs and visual sensors, it is not feasible to directly calculate the linear transformation relationship between the two [15]. Therefore, a new stereo matching method has been proposed in the study. In the inertial visual matching process, the main camera and ISE are first activated, and the left or right wrist is moved closely against a straight line in the environment. The collected data will first undergo Kalman filtering for denoising, and then use the least-squares method to fit these coordinate points onto a straight line. The representation of the line to be fitted is shown in the following equation:

$$\begin{cases} x = x_0 + mt \\ y = y_0 + nt \\ z = z_0 + kt \end{cases}. \quad (1)$$

In Eq. (1), x_0, y_0, and z_0 represent the mean coordinates after filtering. The equation to be solved can be transformed into the following equation:

$$C(t) = P + Dt. \tag{2}$$

In Eq. (2), D represents the direction vector of the corresponding line, and P represents the sampling point. The core idea of the least-squares method is to find suitable parameters that minimize the sum of distances from all points to a given line. According to the basic rules of vector operation, this optimization objective is represented as:

$$f = \sum_{i=1}^{N}(V_i^2 - (V_i D)^2). \tag{3}$$

In Eq. (3), V_i represents the vector of point P. The purpose of the least-squares method is to find the minimum value of a function f, which requires taking the derivative of f and finding the point where the derivative equals zero. The result of taking the derivative of f is shown in the following equation:

$$\frac{\partial f}{\partial P} = 2(I - DD^T)\sum V_i = 0 \tag{4}$$

In Eq. (4), I represents the identity matrix. Given that P is deterministic and the modulus of D is 1, the result of multiplying the transposes of D and D is 1. Therefore, f can be expressed as:

$$f = D^T SD \tag{5}$$

The calculation of S in Eq. (5) is shown in the following equation:

$$S = \sum((V_i^T V_i)I - V_i V_i^T) \tag{6}$$

Eq. (6) is used to find the direction vector of the straight line in the coordinate systems corresponding to the ISE and the main camera. It assumes that the coordinate system of the ISE is moved to the origin position of the corrected coordinate system of the main camera. For the convenience of subsequent calculations, the study normalized two directional vectors on the same line [16]. These two vectors are connected at one end and can be considered as a common corresponding point. On the grounds of this corresponding point, the rotation matrix can be calculated. It assumes that the rotation matrix is R, the translation matrix is T, and the sample center of a certain line on the ISE is N. In the corrected main camera plane, the analytical expression of this line is shown in the following equation:

$$z = ax + by + c. \tag{7}$$

After joint analysis, Eq. (8) can be obtained:

$$[a\, b\ -1]T = R[a\, b\ -1]N + c \tag{8}$$

It assumes that N pairs of corresponding lines were ultimately collected, and three different pairs of corresponding lines can be selected each time. To ensure the stability of the results, each sampling result should be different [17]. Each set of data can calculate a set of T values. Finally, it takes the average of these $3N$ sets of calculation results to achieve coordinate alignment between the ISE and the visual sensor. As the sampling period of the ISE is only 0.04 s, the three visual machines are divided into a total of 15 frames per second. In the initial stage, the error of the ISE is small and there is no frame loss problem, so it is chosen as the reference system for the federated Kalman filter. Considering the characteristics of the federated Kalman filter, its update process does not require all sensors to provide data. Using the sampling period of the ISE as the main filter period, even if three visual sensors do not provide data within a certain period, the data fusion process can still proceed in an orderly manner. In data fusion, both ISEs and visual sensors have time-varying measurement errors. To ensure the adaptability of the system, this study calculates the function between the error of the two and time, and utilizes time-varying allocation coefficients. According to the principle of information conservation, the cumulative error covariance matrix of the ISE is normalized. The data fusion algorithm is shown in Figure 2. In Figure 2, IMU represents the ISE, which can detect the acceleration and angular velocity in real time, providing accurate dynamic data for fast actions such as turning. IMU sensors can be attached to joints such as wrists, ankles, or waist to capture subtle motion changes and compensate for blind spots that visual sensors cannot capture at certain angles. Due to the difference in joint coordinates collected by ISE and visual sensors, it is not feasible to directly calculate the linear transformation relationship between the two. The proposed stereo matching method involves collecting data by moving the wrist along a straight line. After denoising through Kalman filtering, the line is fitted using the least-squares method, and the direction vector is calculated and normalized. Based on the corresponding points, the rotation and translation matrices are solved to ultimately achieve coordinate alignment. This data fusion method is innovative and unique in the field of Tai Chi action recognition.

3.2 Tai Chi movement segmentation and recognition on the grounds of the DBSCAN algorithm

When dealing with the problem of recognizing Tai Chi movements, it is first necessary to preprocess the continuous trajectory of Tai Chi movements. To facilitate subsequent feature extraction and classification, this study rotates all data to be in the same initial position [18]. Then, using the

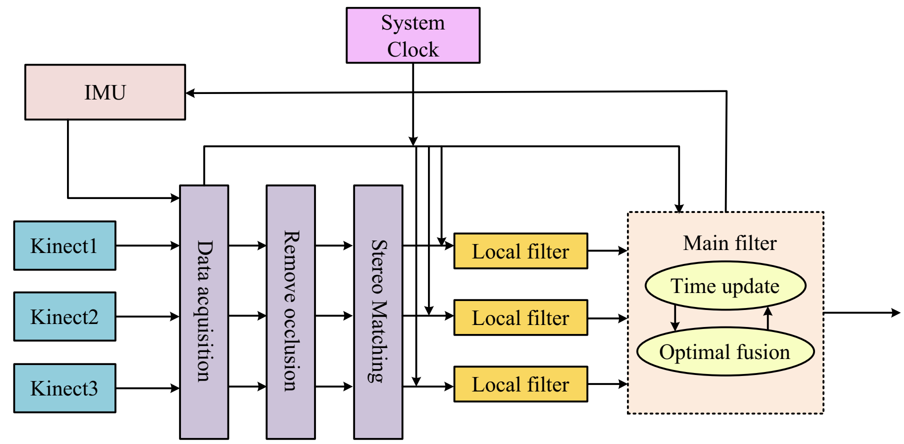

Figure 2: Data fusion algorithm flow.

MDL principle, key feature points are accurately extracted from these continuous trajectories. On the grounds of these feature points, it cuts the original trajectory into multiple subtrajectories and further converts these subtrajectories into four-dimensional feature vectors. When solving classification problems, this study chooses the DBSCAN algorithm, which has the advantage of not requiring to know the number of clusters in advance and can effectively discover clusters of any shape. To achieve the best clustering effect, this study optimizes the clustering parameters by combining the statistical characteristics of the samples and two classic clustering evaluation indicators. The trajectory segmentation and clustering process is shown in Figure 3.

The effectiveness of trajectory segmentation largely depends on the quality of the selected feature points [19]. To ensure optimal segmentation results, feature points should capture the most significant changes in the trajectory as much as possible. It assumes that the trajectory G consists of a series of points p_1, p_2, $p_3 ... p_n$, and the selected feature points are p_{c1}, p_{c2}, $p_{c3} ... p_{c\mathrm{par}}$. By using these feature points, the entire trajectory is divided into multiple subtrajectories. By sequentially connecting these feature points, an approximate representation of the original trajectory is obtained. In the MDL principle, $L(H)$ serves as the sum of the lengths of all subtrajectories, while $L(D|H)$ serves as the sum of the differences between all subtrajectories and the original trajectory. The calculation of $L(H)$ is shown in the following equation:

$$L(H) = \sum_{j}^{\mathrm{par}_i - 1} \log_2(\mathrm{len}(p_{c_j}, p_{c_{j+1}})). \tag{9}$$

In Eq. (9), $\mathrm{len}(p_{c_j}, p_{c_{j+1}})$ represents the Euclidean distance. The calculation of $L(D|H)$ is shown in the following equation:

$$L(D|H) = \sum_{j=1}^{\mathrm{par}_i - 1} \sum_{k=c_j}^{c_{j+1}-1} \log 2(d_\perp(p_{c_j} p_{c_{j+1}}, p_k p_{k+1}) d_\theta \times (p_{c_j} p_{c_{j+1}}, p_k p_{k+1})). \tag{10}$$

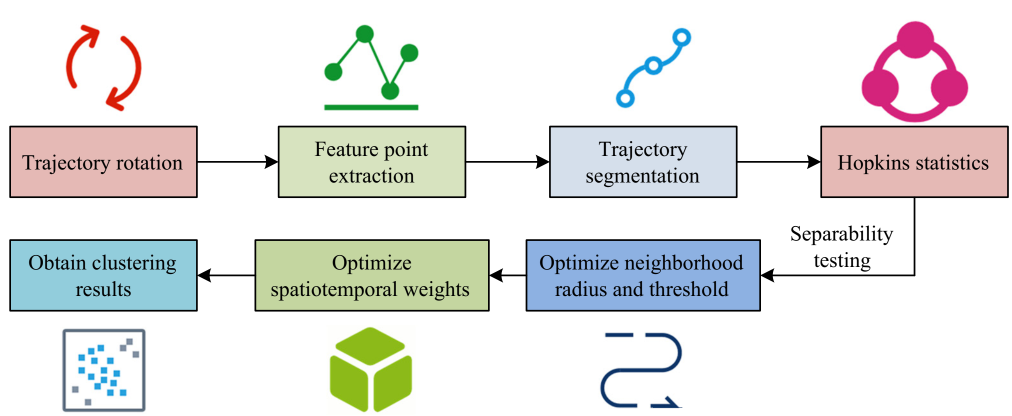

Figure 3: Track segmentation and clustering process.

In Eq. (10), $d_\perp$ represents the vertical length of two trajectory segments L_i and L_j with the same dimension. d_θ serves as the angular distance between the two line segments. Logarithmic functions are used to compress larger values into smaller ranges, avoiding computational problems caused by excessively large values. At the same time, logarithmic functions transform data that originally showed exponential growth into a more easily analyzable form, facilitating subsequent processing and analysis of logarithms. Among them, the calculation of $d_\perp$ is shown in the following equation:

$$d_\perp(L_i, L_j) = \frac{L_{\perp 1}^2 + L_{\perp 2}^2}{L_{\perp 1} + L_{\perp 2}} \tag{11}$$

In Eq. (11), $L_{\perp 1}$ serves as the Euclidean distance between the starting point of two line segments and the projection point, and $L_{\perp 2}$ represents the Euclidean distance between the endpoint of two line segments and the projection point. The calculation of d_θ is shown in the following equation:

$$d_\theta(L_i, L_j) = \begin{cases} \|L_j\| \times \sin\theta, & 0 \le \theta < 90^\circ \\ \|L_j\|, & 90^\circ \le \theta < 180^\circ \end{cases} \tag{12}$$

In Eq. (12), $\|L_j\|$ represents the length of L_j. The distance between two trajectories is defined as:

$$\mathrm{dist}(L_i, L_j) = \omega_\perp d_\perp(L_i, L_j) + \omega_\| d_\|(L_i, L_j) + \omega_\theta d_\theta(L_i, L_j). \tag{13}$$

In Eq. (13), $\omega_\perp$, $\omega_\|$, and ω_θ represent the weights of vertical, parallel, and angular distances, respectively, and $d_\|$ represents the parallel length of two trajectories. A schematic diagram of distance definition is shown in Figure 4.

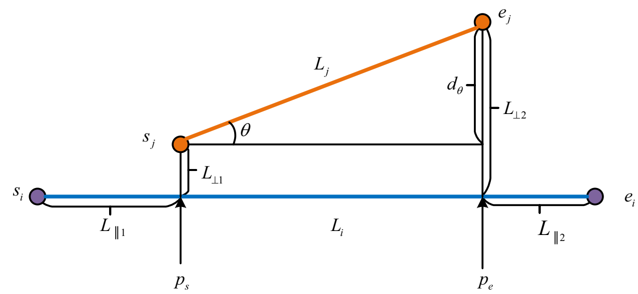

Figure 4: A schematic representation of the definition of the distance.

After determining the correlation function between $L(H)$ and $L(D|H)$, it is then essential for calculating which specific points on a trajectory belong to feature points. This study mainly chooses to search for local optimal solutions to approximate and replace the global optimal solution. By selecting two points p_i and p_j on the original trajectory, two cost functions can be constructed, as represented by the following equation:

$$\begin{cases} \mathrm{MDL}_{\mathrm{par}}(p_i, p_j) = L(H) + L(D|H) \\ \mathrm{MDL}_{\mathrm{nopar}}(p_i, p_j) = L(H) \end{cases}. \tag{14}$$

In Eq. (14), $\mathrm{MDL}_{\mathrm{par}}(p_i, p_j)$ is the encoding length required for the line formed by connecting feature points, while $\mathrm{MDL}_{\mathrm{nopar}}$ is used to measure the encoding length of the original trajectory. If any point p_k between points p_i and p_j satisfies $\mathrm{MDL}_{\mathrm{par}}(p_i, p_k) < \mathrm{MDL}_{\mathrm{nopar}}(p_i, p_k)$, to obtain a local optimal solution, this study considers extending it appropriately outward. In this way, the global optimal solution can be approximated and the points that meet the conditions can be selected as feature points. The obtained exact and approximate solutions are illustrated in Figure 5.

After dividing the entire action trajectory into multiple small segments on the grounds of the position of feature points, the points on the original trajectory are arranged in chronological order, and each subtrajectory has its initial coordinates. If clustering subtrajectories directly, it may lead to excessive redundant information, thereby reducing the accuracy of trajectory classification [20]. Therefore, it is necessary to vectorize the subtrajectories to retain their directional information and convert the original trajectory into a series of connected vectors to obtain a point cloud structure distributed around the origin. Meanwhile, it is necessary to retain the quantity of sampling points contained in each subtrajectory and calculate the time corresponding to the subtrajectory on the grounds of the number of sampling points. Traditional clustering methods require a predetermined number of classifications, so they are not applicable. DBSCAN, on the other hand, can automatically determine the final number of classifications from the data. Therefore, a trajectory clustering algorithm on the grounds of an improved spatial-temporal distance-based DBSCAN Algorithm (IST DBSCAN) was adopted in the study. This method does not require presetting the number of categories, which can better adapt to actual situations and improve the accuracy and reliability of clustering. Meanwhile, the algorithm replaces the Euclidean distance

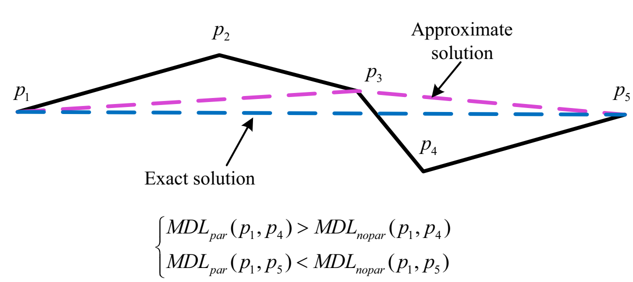

Figure 5: A schematic representation of the exact and approximate solutions.

with the spatiotemporal distance, which better reflects the true distribution of trajectories. In clustering, it can automatically determine the number of clusters, handle noise, and discover clusters of any shape. By determining the distance between the point and core point, it determines whether it is a core or boundary point, and then performs clustering. The advantage is that it does not require a preset number of clusters and is robust to outliers, improving the accuracy and reliability of clustering. In addition, it considers the limitations of IST-DBSCAN in dealing with environmental changes, including changes in lighting, background noise, and inaccurate sensors. To this end, the study further introduced particle filters, which track trajectories through a set of sample points during data collection and can better adapt to non-linear trajectory data, improving the effectiveness of handling complex situations. During the operation, a set of particles is first initialized and distributed based on prior information or preliminary observations. Next, based on the known state transition model, each particle is predicted to estimate the state of the object at the next moment. Then, the particle filter calculates the weight of each particle based on the observed data, which reflects the degree of matching between the particle and actual observation, and adjusts the distribution of the particles through a resampling step. Through this series of steps, particle filters can effectively respond to nonlinear changes in dynamic environments, improve data accuracy, and provide smoother and more accurate trajectory data for IST-DBSCAN, enhancing the clustering performance of the algorithm in complex environments.

To further enhance the user experience of the Tai Chi action recognition system, a user interface and mobile application for Tai Chi coaches and learners have been developed. This interface has a clear navigation design to meet the needs of different users. For non-technical users, an intuitive layout and actionable features are adopted to reduce the learning curve, allowing users to quickly get started and obtain the required feedback through simple operations. At the same time, the system combines a real-time visual feedback function, which can provide real-time feedback on the accuracy, fluency, and posture correctness of users' movements during Tai Chi training. When the learner's action deviates, the system can indicate the incorrect part of the action through a graphical interface and remind the user with color illustrations. In addition, for users of different levels, the system can provide a personalized progress tracking function, allowing learners to view their performance changes in multiple training cycles, thereby improving their skills more targetedly.

4 Analysis of Tai Chi movement segmentation and recognition on the grounds of MSDF and DBSCAN algorithm

This section first verifies the effectiveness of the MSDF algorithm, the MDL principle, and the IST-DBSCAN clustering algorithm. Subsequently, the effectiveness of the Tai Chi action recognition method on the grounds of the MSDF-DBSCAN algorithm was verified. It selected ten unique moves as samples for testing for verifying the recognition superiority.

4.1 Experimental analysis of MSDF and DBSCAN clustering algorithm

To verify the effectiveness of the Tai Chi movement segmentation and recognition algorithm proposed by the research institute, experimental analysis was conducted. The experiment recruited 100 professional athletes and enthusiasts with different levels of Tai Chi practice as research subjects. During the experiment, three Kinect sensors were used to form a triangular layout. Using multi-threaded tools to achieve synchronous sampling of three cameras and real-time recording of the correspondence between each sampling point and system time. At the same time, ISEs are attached to joints such as the wrists, ankles, or waist of the body to detect acceleration and angular velocity in real time. Pressure sensors are installed on the soles of the feet to detect the contact force between the body and the ground. Through the collaborative work of these sensors, motion data during Tai Chi practice is collected as an experimental dataset. The study first validated the effectiveness of the MSDF algorithm. This experiment took the single whip trajectory of the left wrist as an example to obtain its projection on the x–y plane of the world coordinate system. The experimental environment is shown in Table 1.

The MSDF results of a single whip action are shown in Figure 6. As shown in Figure 6, the fused multi-sensor data can more accurately reflect the projection of the single-whip movement of the left wrist on the x–y plane of the world coordinate system. Compared with single-sensor data, the fused data can better capture the details and changes of actions, thereby improving the accuracy and reliability of action recognition. This proves the effectiveness of the MSDF algorithm in action recognition.

Table 1: Experimental environment setting

Experimental	Parameter
System	Windows 10.0
Running platform	NVIDIA GeForce GTX 1080 Ti GPU
Integrated development environment	Pycharm
Deep learning framework	PyTorch
Programming language	Python 2.7

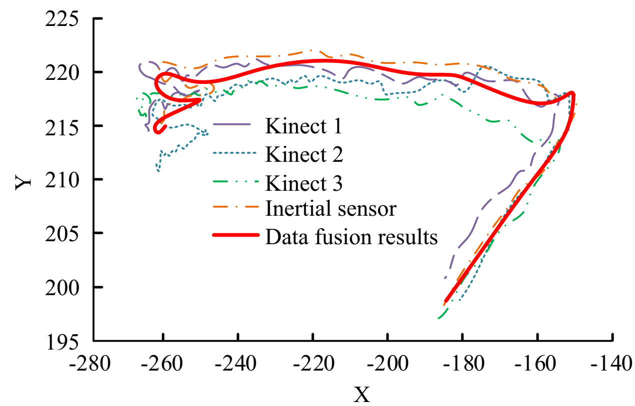

Figure 6: MSDF results for the single-whip action.

To verify the effectiveness of the MDL principle in selecting feature points for Tai Chi movement trajectories, the left wrist joint trajectory of the move Yunshou was studied as the experimental object. The feature point selection results of the left wrist trajectory of the action cloud hand and the representation curve of the feature points are shown in Figure 7. In Figure 7(a), the red dots represent the feature point positions calculated using the MDL principle. The remaining areas represent points that have not utilized the MDL principle. Figure 7(b) shows the trajectory curve corresponding to the original data. Figure 7(c) represents the trajectory curve formed by connecting feature points. Figure 7 shows that the trajectory curve formed by connecting the feature points calculated using the MDL principle is no different from the original trajectory, indicating that the MDL principle can accurately describe the trajectory of Tai Chi movements.

Before evaluating the clustering quality, it is necessary to verify whether there are potential clusterable structures in the original Tai Chi dataset. The research mainly selects

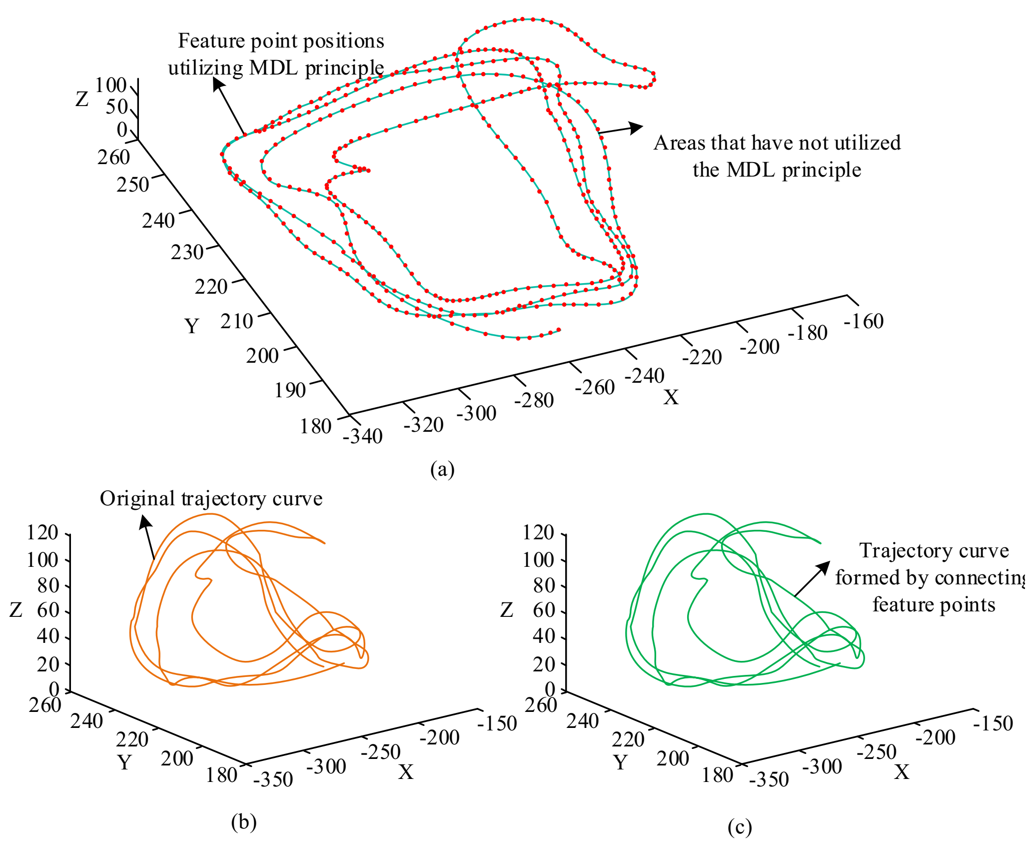

Figure 7: The result of feature points of the left wrist trajectory of the cloud hand and the representation curve of feature points: (a) Feature point selection results, (b) corresponding curve of raw data, and (c) corresponding curve of feature points extracted by MDL principle.

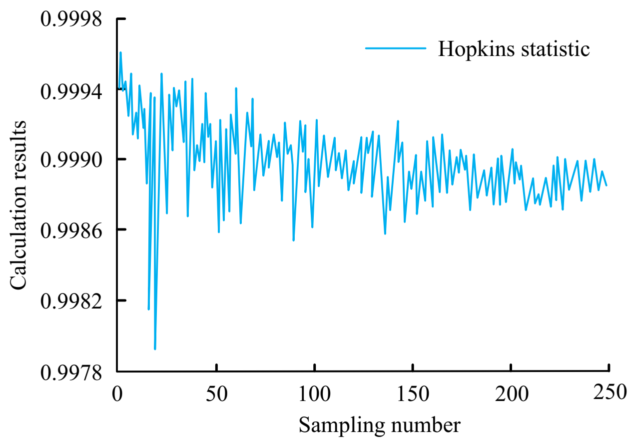

Figure 8: Trajectories of the Hopkins statistics with the number of samples sampled.

Hopkins statistics to analyze the clusterability of the dataset. When the data in the dataset are evenly distributed, the value of the Hopkins statistic is approximately 0.5. If the value is less than 0.5, it can be considered that the dataset cannot be clustered. The closer the value is to 1, the more excellent the clustering effect of the dataset, indicating better separability of the dataset. Taking Tai Chi Cloud Hand as an example, when assigning equal weights to time distance and space distance in the formula for calculating distance, the trajectory of the Hopkins statistic with respect to the number of samples is shown in Figure 8. Figure 8 shows that the value of the Hopkins statistic fluctuates within the range of [0.9979, 0.9996], and its value is very close to 1, which strongly proves the good separability of the sample.

After determining the separability of the data, the effectiveness of IST-DBSCAN clustering can be further evaluated. The effectiveness evaluation of clustering on the grounds of internal standards mainly utilizes the inter-class distance and intra-class distance obtained from classification to comprehensively evaluate the clustering effect. Meanwhile, this study compares *K*-means clustering and means shift clustering, and uses the Silhouette Index (DI) for evaluating the clustering effect. The larger the DI calculation result, the more excellent the clustering effect. The experiment mainly utilizes the trajectories of the left wrist joint and the right wrist joint for action recognition. Therefore, the trajectories of these two are separately clustered. First, it sets the step size of the weight change to 1 and gradually increases it from 1 to 99. The correspondence between the DI results and the spatial distance weight ω_d for the joint points of the left and right wrists is shown in Figure 9. Figure 9(a) shows that when clustering trajectories on the left wrist, IST-DBSCAN can achieve the best clustering effect when the value of ω_d is 68, with a DI value as high as 0.42. The highest DI values of *K*-means and means shift clustering methods are only 0.28 and 0.26, respectively. As shown in Figure 9(b), when clustering the trajectories of the right wrist, the clustering effect is best when the value of ω_d is 77, with a DI value of 0.53, which is significantly higher than the other two clustering methods. This indicates that the IST-DBSCAN clustering method has significant clustering effects.

4.2 Experimental analysis of Tai Chi movement recognition

Aiming at verifying the effectiveness of the motion recognition method for Tai Chi on the grounds of the MSDF-DBSCAN algorithm, this study selects ten unique moves

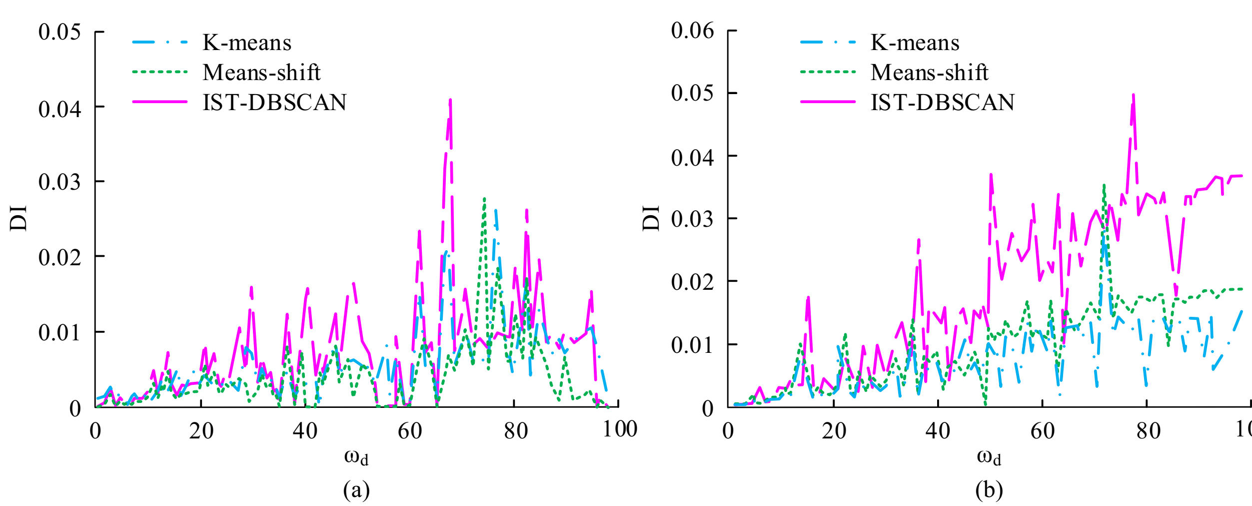

Figure 9: Correspondence between the DI results and the spatial distance weights ω_d: (a) Cluster results of left wrist trajectory and (b) right wrist trajectory.

Table 2: Recognition accuracy of binary classification when training samples account for 25%

	A1 (%)	A2 (%)	A3 (%)	A4 (%)	A5 (%)	A6 (%)	A7 (%)	A8 (%)	A9 (%)	A10 (%)
A10	92.61	90.85	93.35	93.35	92.15	91.36	93.76	91.22	90.97	—
A9	94.53	94.70	92.30	92.30	92.43	94.71	93.48	92.94	—	90.97
A8	90.82	94.56	90.01	92.17	94.14	92.81	9251	—	92.94	91.22
A7	91.75	90.35	93.55	90.85	92.35	89.88	—	92.51	93.48	93.76
A6	90.51	92.76	91.22	93.52	89.97	—	89.88	92.81	94.71	91.36
A5	94.27	92.76	94.35	93.35	—	89.97	92.35	94.14	9243	92.15
A4	94.36	91.17	90.45	—	93.35	93.52	90.85	92.17	92.30	93.35
A3	91.23	91.39	—	90.45	94.35	91.22	93.55	90.01	92.30	93.35
A2	90.99	—	91.39	91.17	92.76	92.76	90.35	94.56	94.70	90.85
A1	—	90.99	91.23	94.36	94.27	90.51	91.75	90.82	94.53	92.61

as samples for classification, prioritizing the moves that appear more frequently in Wu-style Tai Chi. The final ten action samples determined for classification include A1 Grasping Sparrow Tail, A2 Single Whip, A3 Cross Hand, A4 Cloud Hand, A5 Underwater Needle, A6 Flash Back, A7 Oblique Flying, A8 Overturning Monkey, A9 Tai Chi Qi Shi, and A10 Moving and Blocking Hammer. The study separately extracts and converts each move into Fisher vectors, and divides the sample data into two parts: training samples and testing samples. The training samples are utilized for training and optimizing the classifier, while the testing samples are utilized for evaluating the performance of the classifier. Considering the issue of limited training samples, the study adopted transfer learning methods to increase the sample size. Specifically, the study fine-tunes pretrained deep neural network models to adapt to specific movements and styles of Tai Chi. The fine-tuning process focuses on the unique movement characteristics of each Tai Chi style, optimizing the last few layers of the network to enable the model to recognize different styles of moves while avoiding overfitting and improving training efficiency. The proportion of training samples possesses a direct influence on the accuracy of the classifier. The experiment mainly utilizes the motion trajectory of the left hand for single-joint-based action recognition. The accuracy of the classifier is calculated in binary and multiclassification scenarios when the training samples accounted for 25 and 75%, respectively. The recognition accuracy of binary classification when the training sample accounts for 25% is shown in Table 2. According to Table 2, with 25% of the training samples, the highest accuracy of the classifier is only 94.71%, and the lowest accuracy is only 89.88%.

The recognition accuracy of binary classification when the training sample accounts for 75% is shown in Table 3. According to Table 3, with 75% of the training samples, the classifier achieves the highest accuracy of 99.80% and the lowest accuracy of 97.28%. Relative to the situation where the training sample is 25%, its highest and lowest accuracies have improved by 5.09 and 7.32%, respectively. This indicates that the higher the training sample, the higher the accuracy of the classifier. In addition, it can be seen that there are significant differences in the classification accuracy of certain moves under different training sample ratios. Among them, the accuracy of the A5 underwater needle is lower when the training sample accounts for 25%, but significantly improves when the training sample

Table 3: Recognition accuracy of binary classification when training samples account for 75%

	A1 (%)	A2 (%)	A3 (%)	A4 (%)	A5 (%)	A6 (%)	A7 (%)	A8 (%)	A9 (%)	A10 (%)
A10	97.94	99.17	98.74	98.37	99.41	98.47	99.67	99.26	99.80	—
A9	98.39	99.68	99.26	98.26	98.28	99.13	99.26	98.67	—	99.80
A8	97.48	98.14	98.28	99.27	99.34	98.57	99.58	—	98.67	99.26
A7	98.95	99.28	97.91	98.34	98.65	99.38	—	99.58	99.26	99.67
A6	97.64	98.69	98.28	97.51	98.47	—	99.38	98.57	99.13	98.47
A5	99.17	98.14	99.05	98.63	—	98.47	98.65	99.34	98.28	99.41
A4	97.64	98.18	97.28	—	98.63	97.51	98.34	99.27	98.26	98.37
A3	98.54	97.85	—	97.28	99.05	98.28	97.91	98.28	99.26	98.74
A2	98.68	—	97.85	98.18	98.14	98.69	99.28	98.14	99.68	99.17
A1	—	98.68	98.54	97.64	99.17	97.64	98.95	97.48	98.39	97.94

accounts for 75%. This indicates that the A5 underwater needle has a high similarity in motion trajectory with other moves, which makes it easy to be misclassified when there are few training samples. The A3 cross-hand and A4 cloud-hand movements are relatively similar in their form of action, both involving the rotation and movement of the hands, but the specific trajectory and rhythm of the movements are different. Therefore, the classifier finds it difficult to accurately distinguish these subtle differences when there are few training samples, leading to misclassification. Although there are differences in the specific movement trajectory and direction of the hands between the A2 single whip and A7 oblique flying moves, they exhibit similar movement patterns during the extension or retraction stages of the arms, making it difficult for the classifier to accurately distinguish. Therefore, it is necessary to increase training samples to improve classification accuracy.

This study further validates the effectiveness of the proposed action recognition method and compares it with current advanced action recognition algorithms. This includes comparing the recognition performance of the global adaptive graph convolutional network (GAGCN) and hybrid long short-term memory convolutional neural network model (HLSTM-CNN). The recognition accuracy of each algorithm in binary and multi-classification scenarios was studied and calculated. The results of ten tests are shown in Figure 10. As shown in Figure 10 (a), in the case of binary classification, the mean accuracy of the MSDF-DBSCAN algorithm is as high as 98.52%, which is 5.43 and 2.94% higher than that of GAGCN and HLSTM-CNN algorithms, respectively. As shown in Figure 10(b), in the multiclassification scenario, the mean accuracy of the MSDF-DBSCAN algorithm is as high as 92.33%, which is significantly better than other algorithms. The MSDF-DBSCAN algorithm has a more significant action recognition effect.

In order to verify the accuracy and universality of the Tai Chi action recognition method based on the MSDF-DBSCAN algorithm in the styles of Chen style, Yang style, and Sun style Tai Chi, an extended experiment was designed for research. On the basis of the original Wu-style Tai Chi dataset, samples of Chen style, Yang style, and Sun-style Tai Chi movements were added. Each style selects representative moves, and Chen style Tai Chi is characterized by changes in strength and energy, including getting rid of energy, golden chicken independence, single-whip style, turning back and changing steps, and white crane flapping wings. Yang style Tai Chi is known for its smoothness, softness, and slowness, including Tai Chi Qi, Single Whip Style, Cloud Hand, Golden Rooster Independent, and White Crane Bright Wings. Sun style Tai Chi emphasizes rapid changes and clever steps, including pushing hands, lifting knees, hanging flower steps, shaking palms, sinking shoulders, and releasing strength. In a multiclass scenario with 75% of the training samples, the recognition accuracy of each algorithm is shown in Figure 11. As shown in Figure 11, in the Chen style Tai Chi style, the motion recognition accuracy of the algorithm proposed by the research institute is as high as 94.21%, which is 20.63% higher than that of the GAGCN algorithm. In the style of Yang style Tai Chi, the action recognition accuracy of the MSDF-DBSCAN

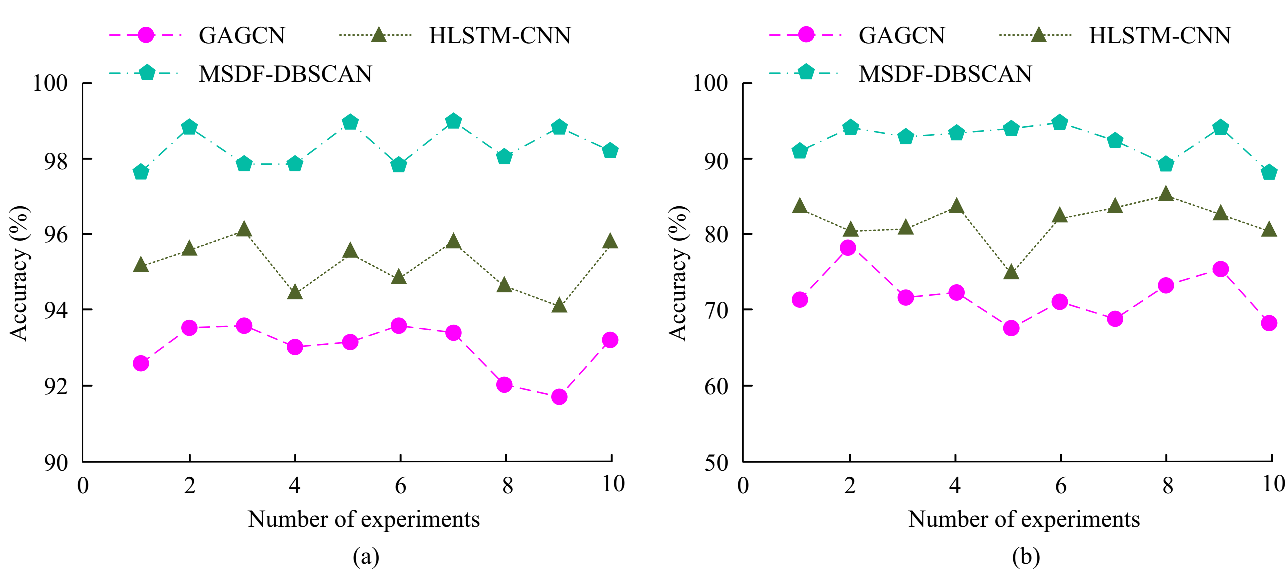

Figure 10: The recognition accuracy of each algorithm in binary and multiclassification scenarios: (a) Accuracy in binary classification scenarios and (b) accuracy in multi classification scenarios.

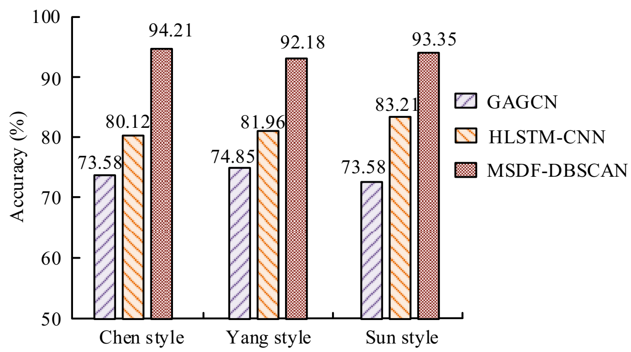

Figure 11: Accuracy of different algorithms in recognizing movements of Chen style, Yang style, and Sun style Tai Chi.

algorithm is as high as 92.18%, which is 17.33% higher than that of the HLSM-CNN algorithm. In the style of Sun style Tai Chi, the action recognition accuracy of the MSDF-DBSCAN algorithm is as high as 93.35%, significantly higher than that of the other two models. This confirms the accuracy and universality of the MSDF-DBSCAN algorithm in recognizing different styles of Tai Chi movements.

The study further validated the recognition efficiency of different algorithms in action recognition tasks. By comparing the performance of the MSDF-DBSCAN, GAGCN, and HLSM-CNN algorithms, the latency and frames per second (FPS) of each algorithm in real-time processing were measured. The real-time processing delay and FPS of each algorithm in different styles of Tai Chi are shown in Table 4. According to Table 4, among all styles of Tai Chi, the delay of MSDF-DBSCAN is lower than other algorithms. The delays of Wu style, Chen style, Yang style, and Sun style are only 55, 52, 51, and 58 ms, respectively. This indicates that the algorithm can process input data more quickly and provide real-time feedback. At the same time, the FPS of the algorithm is 18.5 frames per second, demonstrating extremely high processing power, ensuring the smoothness of the video stream and real-time feedback effect.

The study finally evaluated the effectiveness of the Tai Chi gesture recognition system based on the MSDF-DBSCAN algorithm in long-term use. The study recruited

Table 4: Real-time processing delay and FPS of various algorithms in different styles of Tai Chi

Algorithm	Delay (ms)				FPS (frames per second)
	Wu style	Chen style	Yang style	Sun style	
MSDF-DBSCAN	55	52	51	58	18.5
GAGCN	95	95	97	98	12.7
HLSTM-CNN	90	84	83	88	14.3

100 Tai Chi learners and randomly divided them into an experimental group and a control group, with 50 students in each group. The experimental group used a Tai Chi action recognition system based on the MSDF-DBSCAN algorithm for Tai Chi teaching. The control group used traditional Tai Chi teaching methods. The teaching frequency is three times a week, with 60 min of training each time. Before the experiment begins, a preliminary test is conducted on the basic movements of Tai Chi for all participants, and the accuracy and fluency of each student's movements are recorded. At the end of each month, a mid-term evaluation of the students is conducted to record their accuracy and fluency in movements. The accuracy and fluency scores of the students' movements after 3 months are recorded. The total score for fluency is 50 points, evaluated by 10 experts, and the average is taken as the final score for each learner. The experimental results are shown in Table 5. According to Table 5, in terms of action accuracy, the experimental group had an accuracy rate of 88.5% before the start of the experiment. One month later, the accuracy rate increased to 95.3%, reaching 97.8% after 2 months, and further improving to 98.8% by 3 months. This indicates that the system based on the MSDF-DBSCAN algorithm can provide real-time feedback during long-term training, helping learners correct action errors in a timely manner. Compared to the control group, the improvement in action accuracy was not significant, with only a 2.2% increase. In terms of motor fluency, the experimental group's fluency score increased from an initial 35.2

Table 5: Accuracy and fluency of movements

	Accuracy (%)		Fluency (points)	
	Experimental group	Control group	Experimental group	Control group
Before the experiment	88.5	87.9	35.2	35.7
1 month later	95.3	88.0	44.8	37.2
2 months later	97.8	88.2	48.2	38.5
3 months later	98.8	90.1	48.6	40.2

points to 48.6 points, which has a long-term effect on improving motor fluency. The fluency score of the control group increased from 35.7 points to 40.2 points, with a relatively small improvement, which is inferior to the teaching method proposed in the study.

5 Conclusion

The recognition of movements in Tai Chi helps athletes understand whether their movements are correct, thereby improving their technical level. On the grounds of this, the study introduced a Tai Chi action recognition method on the grounds of the MSDF-DBSCAN algorithm. The results showed that compared with single-sensor data, the fused data can better capture the details and changes of actions, thereby improving the accuracy and reliability of action recognition. Meanwhile, the trajectory curve formed by connecting the feature points calculated using the MDL principle was indistinguishable from the original trajectory. In the experiment of analyzing the clusterability of the dataset, the value of the Hopkins statistic fluctuated within the range of [0.9979, 0.9996], with a value close to 1. This indicates that the dataset has good clusterability. For action recognition based on the IST-DBSCAN algorithm, data with high clusterability can enable the algorithm to more accurately identify different action patterns, reducing the impact of noise and interference. When clustering trajectories on the left wrist, DBSCAN's DI value was as high as 0.42. When clustering trajectories on the opponent's right wrist, DBSCAN's DI value was as high as 0.53, significantly higher than the other two clustering methods. This algorithm is capable of achieving precise classification of wrist trajectory data.

In addition, with 25% of the training samples, the classifier on the grounds of the MSDF-DBSCAN algorithm had the highest accuracy of only 94.71% and the lowest accuracy of only 89.88%. With 75% of the training samples, the classifier reached a maximum accuracy of 99.80% and a minimum accuracy of 97.28%. In the style of Chen-style Tai Chi, the motion recognition accuracy of the algorithm proposed by the research institute is as high as 94.21%, which is 20.63% higher than the GAGCN algorithm. In the style of Yang style Tai Chi, the action recognition accuracy of the MSDF-DBSCAN algorithm is as high as 92.18%, which is 17.33% higher than that of the HLSM-CNN algorithm. In the style of Sun-style Tai Chi, the action recognition accuracy of the MSDF-DBSCAN algorithm is as high as 93.35%, significantly higher than the other two models. The delay of the MSDF-DBSCAN algorithm in Wu style, Chen style, Yang style, and Sun style Tai Chi is only 55, 52, 51, and 58 ms, respectively, significantly lower than other algorithms. There are differences in movement trajectory, rhythm, and posture among different styles of Tai Chi movements, and the MSDF-DBSCAN algorithm can adapt to these differences and has good generalization ability, indicating that this algorithm has broad application prospects in the field of Tai Chi movement recognition. However, the Tai Chi moves involved in the experiment are not sufficient to cover all factions of Tai Chi moves so that subsequent research can create new datasets for different Tai Chi teaching objectives.

Funding information: Authors state no funding involved.

Author contributions: Zhehua Fan: study design, data collection, statistical analysis, visualization, writing, and revision of the original draft. Fengqun Huang: data collection, statistical analysis, and revision of the manuscript. All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Conflict of interest: Authors state no conflict of interest.

Data availability statement: All data generated or analyzed during this study are included in this published article.

References

[1] Wei Z, Gao B, Liu H, An L, Wang P, Fei C. Assessment algorithm of power supply reliability for customers in rural healthcare centres. J Comm Biotechnol. 2022;27(5):239–52.
[2] Farag W. Kalman-filter-based sensor fusion applied to road-objects detection and tracking for autonomous vehicles. Proc Inst Mech Eng Part I J Syst Control Eng. 2021;235(7):1125–38.
[3] Singh A, Gupta MM. Convergence of machine learning and statistics to predict covid-19 evolution. Acta Inf Malays. 2022;6(1):34–8. doi: 10.26480/aim.01.2022.34.38.
[4] Xu L, Wang Q, Lin X, Yuan L, Ma X. Skeleton-based Tai Chi action segmentation using trajectory primitives and content. Neural Comput Appl. 2023;35(13):9549–66.
[5] Mulcahy A, Holland B, Gosselin K, Pittman A. The use of tai-chi to reduce anxiety among nursing students undergoing simulation. Nurs Educ Perspect. 2020;41(3):183–4.
[6] Yan ZW, Yang Z, Yang J, Chen YF, Zhang XB, Song CL. Tai Chi for spatiotemporal gait features and dynamic balancing capacity in elderly female patients with non-specific low back pain: a six-week randomized controlled trial. J Back Musculoskelet Rehabil. 2022;35(6):1311–9.
[7] Liu H, Gong H, Chen P, Zhang L, Cen H, Fan Y. Biomechanical effects of typical lower limb movements of Chen-style Tai Chi on knee joint. Med Biol Eng Comput. 2023;61(11):3087–101.

[8] Chai W, Li C, Li Q. Multi-sensor fusion-based indoor single-track semantic map construction and localization. IEEE Sens J. 2022;23(3):2470–80.
[9] Sengupta A, Yoshizawa A, Cao S. Automatic radar-camera dataset generation for sensor-fusion applications. IEEE Robot Autom Lett. 2022;7(2):2875–82.
[10] Li J, Zhang Y, Liu X, Zhang X, Bai R. Obstacle detection and tracking algorithm based on multi-lidar fusion in urban environment. IET Intell Transp Syst. 2021;15(11):1372–87.
[11] Purohit M, Singh M, Yadav S, Singh AK, Kumar A, Kaushik BK. Multi-sensor surveillance system based on integrated video analytics. IEEE Sens J. 2021;22(11):10207–22.
[12] Zhang S, Zhao Y, Nguyen DT, Xu R, Alshurafa N. Necksense: A multi-sensor necklace for detecting eating activities in free-living conditions. Proc ACM Interact Mob Wearable Ubiquitous Technol. 2020;4(2):1–26.
[13] Wang Y, Jiang K, Wen T, Jiao X, Wijaya B, Miao J, et al. SGFNet: segmentation guided fusion network for 3D object detection. IEEE Robot Autom Lett. 2023;8(12):8239–46.
[14] Le VH. Deep learning-based for human segmentation and tracking, 3D human pose estimation and action recognition on monocular video of MADS dataset. Multimed Tools Appl. 2023;82(14):20771–818.
[15] Bhosle K, Musande V. Evaluation of Deep Learning CNN Model for Recognition of Devanagari Digit. Artif Intell Appl. 2023;1(2):114–8.
[16] Feng X, Lu X, Si X. Interactive visualization of forearm internal rotation angle in Tai Chi training. J Electron Imaging. 2022;31(6):063061.
[17] Bennett BC. The Somatic Work of Thomas Hanna, Tai Chi, and Kinesiology. Kinesiol Rev. 2020;9(3):236–44.
[18] Chu JT, Szu HH. Biomechanical aspects of Tai Chi Chuan countermeasure against health threats during spaceflight. MOJ App Bio Biomech. 2020;4(5):118–23.
[19] Park J, Krause-Parello CA, Barnes CM. A narrative review of movement-based mind-body interventions: effects of yoga, tai chi, and qigong for back pain patients. Holist Nurs Pract. 2020;34(1):3–23.
[20] Tang C, Chen X, Gong J, Occhipinti LG, Gao S. WMNN: wearables-based multi-column neural network for human activity recognition. IEEE J Biomed Health Inf. 2022;27(1):339–50.