Multi-Sensor Data Fusion Analysis for Tai Chi Action Recognition

Jingying Ouyang[1], Jisheng Zhang[2*], Yuxin Zhao[3], Changhuo Yang[4]
College of Physical Education, Hunan Normal University, Changsha, 410012, China[1, 2]
School of Business, Hunan University, Changsha, 410082, China[3]
School of Computer Science and Engineering, Central South University, Changsha, 410083, China[4]

***Abstract*—The continuous development of action recognition technology can capture the decomposition data of Tai Chi movements, provide precise assistance for learners to correct erroneous movements and enhance their interest in practicing Tai Chi. Inertial sensors and human skeletal models are used to collect motion data. Combined with visual sensors, the motion and trajectory of Tai Chi are processed to obtain the relevant coordinate system of the movement trajectory. Then, the inertial sensor and visual sensor are fused for data processing to standardize the human skeleton model, remove noise interference from the collected information, and improve the smoothness performance of movement trajectories, thereby segmenting and clustering Tai Chi movement trajectories. Finally, the support vector machine and dynamic time-warping algorithm are combined to identify and verify the trajectory of Tai Chi movements. According to the results, in the 25%, 50%, and 75% training sample proportions, the lowest recognition accuracy of the Qi Shi movements was 90.87%, 93.53%, and 98.08%, respectively. The optimal recognition accuracy and standard deviation of single nodes in binary classification were 98.48% and 0.47%, respectively. The best recognition accuracy and standard deviation for multi-joint points in binary classification were 99.77% and 0.16%, respectively. This proves the recognition advantages of binary classification and the superiority of data fusion analysis based on multiple sensors, providing a theoretical basis and technical reference for action recognition technology.**

***Keywords*—Inertial sensor; visual sensors; segmentation clustering; support vector machine; dynamic time warping algorithm**

I. Introduction

With the advancement and rapid development of technology, human motion recognition technology has been widely applied in fields such as medicine, sports, and computer science. According to the human body model and the action mechanism, the standardization of actions can be improved. Tai Chi, is a popular health exercise in society. According to action recognition technology, actions can be effectively imitated, improving teaching level and standardization of actions [1]. However, Tai Chi movements are basically smooth and coherent trajectory movements, which have unity in the coordination of whole body and joint movements. Therefore, it is necessary to collect and process information on Tai Chi movements, as well as segment trajectories to improve recognition accuracy. Action recognition technology has achieved some results in decomposing actions, and identifying motion types and motion patterns. Moreover, there is in-depth research and technological development in the collection and recognition algorithms of action information for human skeleton models [2-4]. However, the current methods for recognizing and decomposing human movements lack more accurate trajectory processing in coherent movements. There is a lack of real-time feedback on monitoring physical fitness and correcting movements during human exercise. Therefore, the inertial sensor and visual sensor are combined to collect, process, and fuse data on motion trajectories. The research innovatively uses Support Vector Machine (SVM) and Dynamic Time Warping (DTW) algorithms to identify and process joint points of continuous actions and their trajectories, aiming to provide theoretical and technical references for the clustering effect of trajectory segmentation and the classification accuracy of action recognition. According to research on the recognition methods of Tai Chi movements, it can not only assist athletes in improving their professional technical level, but also promote the diversified development of other sports. Meanwhile, it can also monitor patients' rehabilitation movements, input human motion commands for intelligent living, and provide practical applications and service experiences for fields such as medicine, public health, and human-computer interaction.

The study is structured into five sections. Section II elaborates on the current research results. Section III analyzes and constructs the data processing and fusion of inertial sensors and visual sensors, improving the smoothness of the trajectory. Section IV is to perform cluster analysis and recognition verification on the trajectory of Tai Chi movements. Section V is a narrative summary of the entire study.

II. Related Works

Action recognition technology has been widely applied and developed in various fields based on human motion information. At present, the research on action recognition technology focuses on action data processing and trajectory clustering algorithms. In recent years, scholars have conducted many explorations on action recognition. Sun et al. used depth maps and human motion recognition datasets to validate human motion recognition. Furthermore, the advantages of human motion recognition based on depth maps were derived [5]. In terns of human behavior identification, Luo et al. proposed long-distance data transmission between sensors and servers to improve the real-time processing capacity. The binary neural network was used to evaluate human activity data, so as to improve the efficiency and accuracy of data operation [6]. Liu et al. used an adaptive multi-scale convolutional network to

*Corresponding Author.

analyze the information features of skeleton joint points and their movements for human skeleton action recognition. The accuracy of action recognition data was verified by combining the character segmentation mode and perspective segmentation mode, thereby proving the superiority of adaptive multi-scale graph convolutional networks in human skeleton action recognition [7]. Regarding the three-dimensional reconstruction of human bones, Liu et al. adopted a skin multi-person linear model and skeleton perception implicit function method. It improved the accuracy and detail processing of the human body model, providing a foundation for human model recognition [8].

There have been many research achievements in the field of human motion recognition technology, including sports, home, and intelligent applications. Liu proposed an algorithm that combines principal component analysis and local binary patterns to collect and process motion images for recognizing athletes' throwing movements. Through image segmentation and recognition, the average recognition rate and accuracy of the movements were improved [9]. To solve the motion recognition technology in sports, Host et al. validated the sports dataset using action systems and computer vision to improve the detection accuracy of athlete training [10]. Regarding indoor behavior recognition for the elderly, Song Y et al. utilized wireless fidelity and video feature fusion to establish an indoor human behavior recognition method. They also combined support vector machines to classify and recognize actions, thereby improving the accuracy of action recognition [11]. Cai et al. used feature extraction tools and multi-scale fusion to extract music and dance features in the automatic generation of folk music and dance movements. A sequence network model was constructed to train features, synthesize new dance sequences, and improve the efficiency of automatic music and dance generation and the accuracy of rhythm matching [12]. For the classification technique of action recognition, Chang et al. used time-frequency features of brain signals and event-related potential phenomena. Furthermore, convolutional neural and long short-term memory network models were constructed to classify the collected EEG signals, demonstrating the superiority of classification algorithms [13]. Regarding baseball trajectory recognition, Seo et al. used an adaptive Kalman filter for velocity compensation. By combining sensor fusion and motion characteristic compensation, the three-dimensional trajectory of global coordinates was estimated, thereby improving baseball trajectory recognition and estimation for high-precision detection [14].

To sum up, although domestic and foreign scholars have conducted model construction on action recognition technologies and classification algorithms, they mainly focus on image processing and recognition of smooth and fast human movements to ensure the accuracy of sports athlete training. However, there is still a lack of in-depth research on clustering and segmentation methods for motion trajectories in daily life scenarios, and the sensor devices used for data collection also lack data fusion. At the same time, there is a lack of consideration for the universality of action recognition. There are significant differences between athlete action recognition and elderly behavior activity feature extraction, which cannot provide a good reference for behavioral rehabilitation in the medical field. Therefore, multi-sensor fusion is used to collect Tai Chi movement data, and the trajectory segmentation algorithm is used to ensure the uniform decomposition of Tai Chi movement, thereby improving the recognition and classification of movement trajectories. At the same time, video explanations of Tai Chi and accurate and consistent movement guidance are provided during fitness exercises.

III. System Construction for Data Collection and Data Fusion

Tai Chi is a continuous movement trajectory. The action data are obtained based on inertial sensors and visual sensors. Then the action trajectory data is collected to complete action recognition.

A. *Data Acquisition and Processing Based on Inertial Sensors and Visual Sensors*

The human posture representation is based on the human skeleton model for collecting action data. To obtain real-time action data, inertial sensors are used for convenient wearable devices. The specific structural composition is shown in Fig. 1.

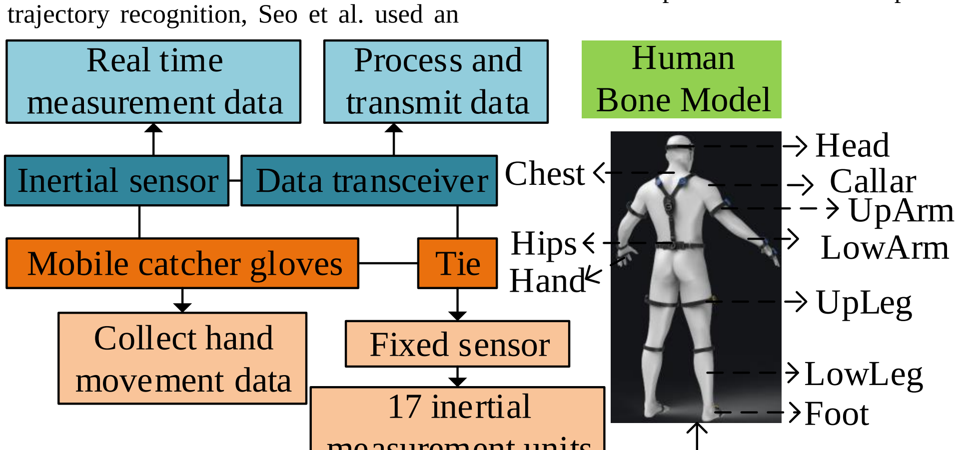


Fig. 1. Data acquisition equipment and its structure based on inertial sensors.

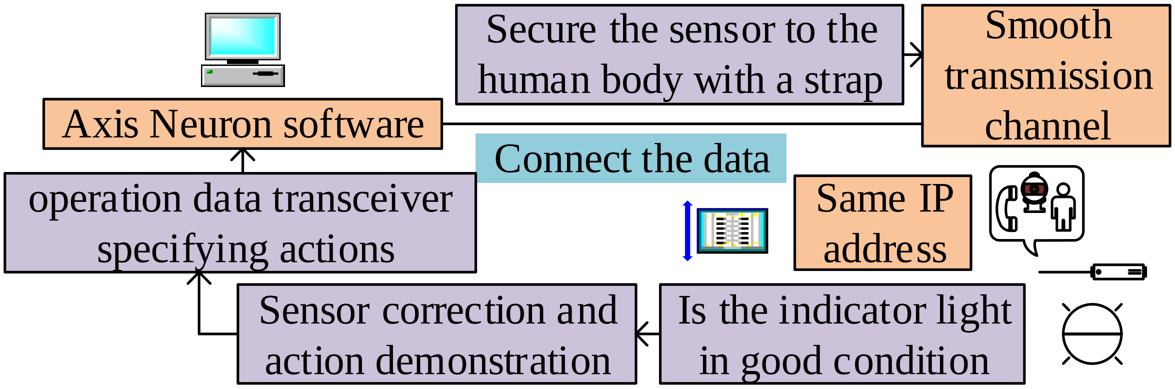


Fig. 2. Schematic diagram of data collection steps.

From Fig. 1, the inertial sensor measures data such as velocity, intensity, and acceleration for each unit. The data transceiver processes the data and synchronously transmits it to the host. The motion capture gloves are only fixed on the hands to obtain data on hand movements. The strap is used to fix the inertial sensor on the human body, completing the data collection process. According to the characteristics of Tai Chi exercise, in addition to the hand movement data of the motion capture gloves, the sensor is fixed in 17 measurement positions on the human body with straps to receive real-time transmission data from the data transceiver. The data collection steps are shown in Fig. 2.

From Fig. 2, during the data collection process, the sensor is fixed to the human body. The connection between the host software and the data transceiver is smooth. Before collecting data, the sensor is corrected and the action is simulated and demonstrated. Finally, the host software is used to operate the data transceiver to complete the data collection of the specified action. BVH, as a data storage format in hierarchical models, is often used in the animation production of human bone models. Initially, each joint point is positioned and its hierarchical relationship is analyzed. Then the motion data of each joint point frame is calculated. The initial coordinate difference between a certain level node and its parent node is expressed as vector form, as shown in Eq. (1).

$$V_0 = \left[x_0, y_0, z_0\right]^T \tag{1}$$

In Eq. (1), x_0, y_0, and z_0 represent the initial offset of the node relative to their child nodes. x, y and z are three directions. In frame $k+1$, the vector of the joint points to the parent node is shown in Eq. (2).

$$V_{k+1} = T_k R_k \left[V_0 \quad 1\right]^T \tag{2}$$

In Eq. (2), T_k is the translation matrix. R_k is the product of rotation matrices in the z, x, and y directions. According to the node transformation relationship and the bone hierarchical structure, the three-dimensional coordinate is shown in Eq. (3).

$$V_{k+1}^n = \left(\prod_{i=0}^{n} T_k R_k^i\right) \times V_0^n \tag{3}$$

In Eq. (3), n represents the number of child node levels. The 0-th level parent node represents the parent node. The difference between the child node and the parent node is $n-1$. Then, the points obtained from the three-dimensional coordinates are connected to obtain the trajectory of Tai Chi movements. In addition, with the rapid development of machine learning, two-dimensional color image models can be extracted from human bone models. The extraction techniques include top-down and bottom-up approaches. The study adopts a bottom-up extraction method, which extracts various body parts or nodes from the image and then pieces them together to form a complete bone model. Openpose, as a bottom-up skeleton recognition method, has a human skeleton model with 18 joints and 25 joints. The study selects a human skeleton model with 25 joints, which includes all joints in BVH format, as well as data extraction for hands and faces. The Openpose algorithm can extract two-dimensional coordinates of human bone models in color images. Corresponding to the position of the depth image, the three-dimensional coordinates of the joint points can be obtained.

The three-dimensional coordinates of depth images can be obtained based on the camera coordinate system. The common depth images include stereo vision method, time-of-flight method, etc. [15]. Based on the coherence and motion range of Tai Chi movements, the KinectV2 somatosensory camera based on the time-of-flight method is taken as the visual sensor for research. Visual sensors cover four coordinate systems in data analysis and calculation. Therefore, its research application in Tai Chi action recognition is shown in Fig. 3.

From Fig. 3, four coordinate systems are obtained, including image coordinate system, pixel coordinate system, camera coordinate system, and world coordinate system. Inertial sensors and stereo sensors are combined to study the relevant coordinate systems of Tai Chi movement trajectory.

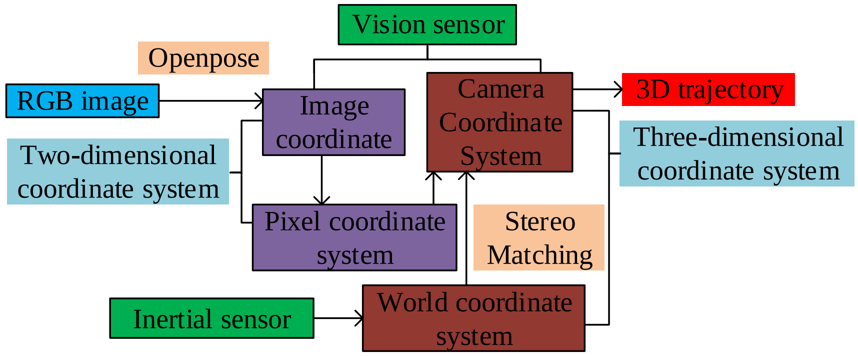


Fig. 3. Coordinate system application and structure diagram of visual sensors.

B. Data Fusion Processing Based on Inertial-Visual Sensors

The imaging captured by a camera is affected by optical properties, which can cause distortion during imaging. Then, based on the camera calibration internal parameter matrix, distortion processing is applied to the depth image and color image. The points of the depth image are corresponding to the color image to obtain a four-channel RGB color mode Depth Map (RGBD) image. Finally, based on the camera coordinate system, the three-dimensional coordinates of the RGBD image are evaluated. For the pixel coordinates of a point in the depth image, the depth value is shown in Eq. (4).

$$\text{Depth value} = u + v$$
$$\begin{cases} u = a_2\left[b_2 d\left(a_1, b_1\right) + \left(1 - b_2\right) d\left(a_1, b_1 + 1\right)\right] \\ v = a_2\left[\left(1 - a_2\right) d\left(a_1 + 1, b_1\right) + \left(1 - b_2\right) d\left(a_1 + 1, b_1 + 1\right)\right] \end{cases} \tag{4}$$

In Eq. (4), (a,b) is the pixel coordinate of a certain point. $a = a_1 + a_2, a_1 \le a$, and $b = b_1 + b_2, b_1 \le b$. Depth value represents the depth value of the point. $d(a,b)$ is the function value of point (a,b) on the depth image. In addition, the depth value of joint points in the three-dimensional coordinates of the depth camera is shown in Eq. (5).

$$\begin{cases} x = D\left(a - a_0\right)\dfrac{1}{f_x} \\ y = D\left(b - b_0\right)\dfrac{1}{f_y} \end{cases} \tag{5}$$

In Eq. (5), (x,y,z) is the three-dimensional coordinate of the joint point. (x,y,z) is its depth value. Finally, the RGBD image is separated to obtain an RGB image with the same size as the depth image. The Openpose algorithm is used to extract pixel coordinates from the joint points of the RGB image and convert them into three-dimensional coordinates. Finally, the visual sensor is used to process data such as noise and errors in three-dimensional coordinates. The image is processed with depth values through guided filters to obtain a single channel binarized image, which is optimized to preserve the distinguishing effect between the edges and interior of the image, thereby clarifying the human body contour information on the image. The morphology of human bones varies. Therefore, the human skeleton model needs to be standardized, so that visual sensors can clearly obtain the three-dimensional coordinates of the human skeleton and ensure the optimization quality of Tai Chi movement trajectory. To remove noise interference from collected data, the study uses Kalman filtering to process motion data and smooth the trajectory. The smoothing indicators, including square integral method, root mean square method, maximum velocity normalized mean method, and maximum velocity normalized mean method are used to test the smoothness and coordination of human motion. The improved Kalman filter enhances the smoothness of motion trajectory. Afterwards, the data collection of the inertial sensor is denoised. The five-point smoothing algorithm is used to optimize its motion trajectory, thereby improving the good smoothness performance of the Tai Chi movement trajectory. The accuracy of inertial sensors is superior to that of visual sensors. Therefore, there are significant errors in the data after the operation. The data fusion method of the two needs to be improved to enhance the quality of data processing. The steps for designing data fusion are shown in Fig. 4.

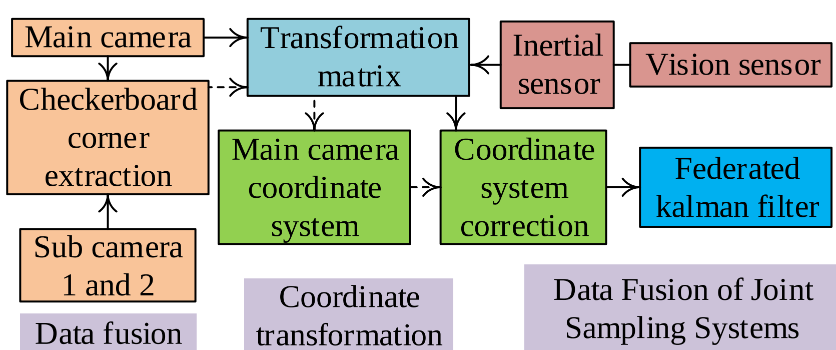


Fig. 4. Data fusion of inertial sensor and visual sensor.

From Fig. 4, the data fusion processing design is formed based on the errors of the two sensors. Three cameras and their coordinate systems are stereo matched. The checkerboard corner extraction algorithm is combined to enhance robustness and reduce errors. Then, coordinate transformation is performed on the inertial visual sensor to accurately identify the three-dimensional coordinates of the human skeleton model. Finally, a joint sampling system is used to fuse the data of the inertial visual sensor. The decentralized multi-sensor data fusion method is adopted, and Kalman filtering is combined to design filters, thereby reducing data errors and efficiently fusing information [16-18]. The main error in inertial sensors is the operational error of gyroscopes and accelerometers. The errors in visual sensors mainly include joint positioning, depth values, and other errors. Therefore, the joint error compensation algorithm and Kalman filtering algorithm are used to process the cumulative error and joint error data, thereby improving the robustness of the fusion system. The stereo matching angle relationship of inertial-visual sensor is shown in Eq. (6).

$$\begin{cases} \alpha = \left| \arccos\left(\dfrac{(n_3)}{\sqrt{n_1^2+n_2^2+n_3^2}} \right) \right| \\ \beta = \left| \arccos\left(\dfrac{(n_1)}{\sqrt{n_1^2+n_2^2+n_3^2}} \right) \right| \\ \alpha+\beta=\varphi+\lambda=\dfrac{\pi}{2} \end{cases} \tag{6}$$

In Eq. (6), α is the angle between z in camera coordinates and the ground plane. β is the angle between x and the ground plane. The converted camera coordinate is (x', y', z'). The angle between z and z' is φ. λ is the angle between x and x'. The information conservation principle of the federated Kalman filtering algorithm is shown in Eq. (7).

$$P=\sum_{i}^{m}\alpha_i+\alpha_m \tag{7}$$

In Eq. (7), P represents the principle of information conservation. The noise variance error is shown in Eq. (8).

$$\begin{cases} \sum_{i=1}^{N}\dfrac{1}{\eta_i}=1 \\ 0\le\dfrac{1}{\eta_i}=\alpha_i\le 1 \end{cases} \tag{8}$$

In Eq. (8), η_i is the measurement noise variance of the sampling system. Finally, the covariance between the inertial sensor and the visual sensor is normalized, as shown in Eq. (9).

$$\alpha_i^k=\frac{\sum_{i=1}^{3}\dfrac{1}{\eta_i}+\dfrac{1}{\eta_m}}{\eta_i} \tag{9}$$

In Eq. (9), η_m is the covariance matrix of the cumulative error for the inertial sensor. Through data fusion design, multiple sensor error analysis and optimal data fusion are completed.

C. Segmentation and Clustering Processing of Tai Chi Movement Trajectories

Tai Chi is a continuous and light movement. To complete the action trajectory recognition, the trajectory needs to be segmented to extract the characteristics of trajectory motion [19-20]. The Minimum Description Length (MDL) is used to extract feature points of a trajectory. The original trajectory is segmented to obtain multiple sub trajectories. Then, the sub-trajectories are transformed into vectors, which are called feature vectors. The lengths of all sub-trajectories during trajectory segmentation are shown in Eq. (10).

$$\begin{cases} M(H)=\sum_{j=1}^{par_i-1}\log_2\left[len\left(p_{c_j},p_{c_{j+1}}\right)\right] \\ M(D|H)=\sum_{j=1}^{par_i-1}\sum_{k=c_j}^{c_{j+1}-1}\log 2\left[d_\perp\left(p_{c_j}p_{c_{j+1}},p_kp_{k+1}\right)\times d_\theta\left(p_{c_j}p_{c_{j+1}},p_kp_{k+1}\right)\right] \end{cases} \tag{10}$$

In Eq. (10), $M(H)$ is the sum of sub-trajectory lengths. p_c is the point on the trajectory. p_{c_j} is the characteristic of the trajectory point. $len\left(p_{c_j},p_{c_{j+1}}\right)$ is the Euclidean distance between p_{c_j} and $p_{c_{j+1}}$. $d_\perp$ and d_θ are the vertical distance and angular distance between trajectory segments, respectively. For the characteristics of trajectory points, it is required to take a local optimal solution, which can replace the global optimal solution. The cost function and feature point conditions are shown in Eq. (11).

$$\begin{cases} MDL_{par}(p_i,p_j)=M(H)+M(D|H) \\ MDL_{nopar}(p_i,p_j)=M(H) \\ MDL_{par}(p_i,p_k)<MDL_{nopar}(p_i,p_j) \end{cases} \tag{11}$$

In Eq. (11), p_i and p_j represent two points on the trajectory, respectively. $MDL_{par}(p_i,p_j)$ is the encoding length function that connects feature points into a line. $MDL_{nopar}(p_i,p_j)$ is the length function of the original trajectory encoding. Due to the small difference between the trajectory and the original trajectory, $M(D|H)$ is zero. Then, the three-dimensional coordinates and time of the sub-trajectories are combined to form a four-dimensional feature vector. The four-dimensional feature vector is classified. Then, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is selected. The algorithm automatically classifies the data according to the set parameters, and then finds clusters of any shape in the dataset and removes noise. The Hopkins statistic is selected to evaluate the clustering trend of the dataset, as shown in Eq. (12).

$$\text{Hopkins statistics} = \frac{\sum y_i}{\sum y_i + \sum x_i} \tag{12}$$

In Eq. (12), $\{x_n\}$ represents uniformly extracting n samples $\{p_n\}$ from the dataset. Then, n sample sets are uniformly extracted from the dataset. The distance between each sample and the nearest sample set in the complement set is $\{y_n\}$. After determining the separability of the dataset, the clustering effect is effectively evaluated. The Davies-Bouldin index (DBI) and Dunn index index (DI) are used to simultaneously evaluate clustering performance. The DBI is shown in Eq. (13).

$$DBI = \frac{1}{k}\sum_{i=1}^{k} \max_{j=1\sim k, j\neq 1}\left(\frac{L_i + L_j}{C_{ij}}\right) \tag{13}$$

In Eq. (13), L_i and L_j represent the average distance between each point in the i-th and j-th clusters and the center of the cluster, respectively. C_{ij} is the center distance between the i-th and j-th clusters. If the value is small, the clustering effect is good. The DI is shown in Eq. (14).

$$DI = \frac{MIN\left\{\min\left\|x_i - y_j\right\|\right\}\begin{cases} MIN, 0 < m \neq n < K \\ \min, \forall x_i \in \Omega_m, \forall y_i \in \Omega_n \end{cases}}{MAX \max\left\|x_i - y_j\right\| \begin{cases} MAX, 0 \leq m \leq K \\ \max, \forall x_i, y_i \in \Omega_m \end{cases}} \tag{14}$$

In Eq. (14), a larger DI indicates better clustering performance. According to the distance evaluation between DBI and DI, the optimal weight can be continuously adjusted. The parameters that need to be adjusted in clustering analysis include spatiotemporal parameters. The spatiotemporal distance is normalized, as shown in Eq. (15).

$$S(W_i, W_j) = w_d \frac{d_\perp(W_i, W_j) + d_\parallel(W_i, W_j) + d_\theta(W_i, W_j)}{\max(d_\perp + d_\parallel + d_\theta)} + \frac{w_t d_t(W_i, W_j)}{\max(d_t)} \tag{15}$$

In Eq. (15), W_i and W_j represent trajectory segments, respectively. $d_\parallel$ is the parallel distance between trajectory segments. w_d is the sum of weights among parallel distance, vertical distance, and angular distance. w_t is the time weight, and $w_d + w_t = 1$. d_t is the time distance. According to the inertial visual sensor and sampling system, the complete Tai Chi movement trajectory is segmented and clustered. Then, its actions are decomposed and recognized. Due to the complexity and diversity of Tai Chi movement trajectories, SVM and DTW are combined to identify and verify single and multiple joint points, respectively. The steps for identifying the two are shown in Fig. 5.

From Fig. 5, based on the specific movements and trajectories of Tai Chi, its movements are identified and tested. SVM is fused to process the trajectory into the same size form, and Fisher vectors are used to normalize the length. The results of DBSCAN are used in a mixed Gaussian model to identify the motion trajectory of a single joint point. For multi-joint recognition verification, DTW is used to optimize the algorithm and directly recognize the segmented actions as templates.

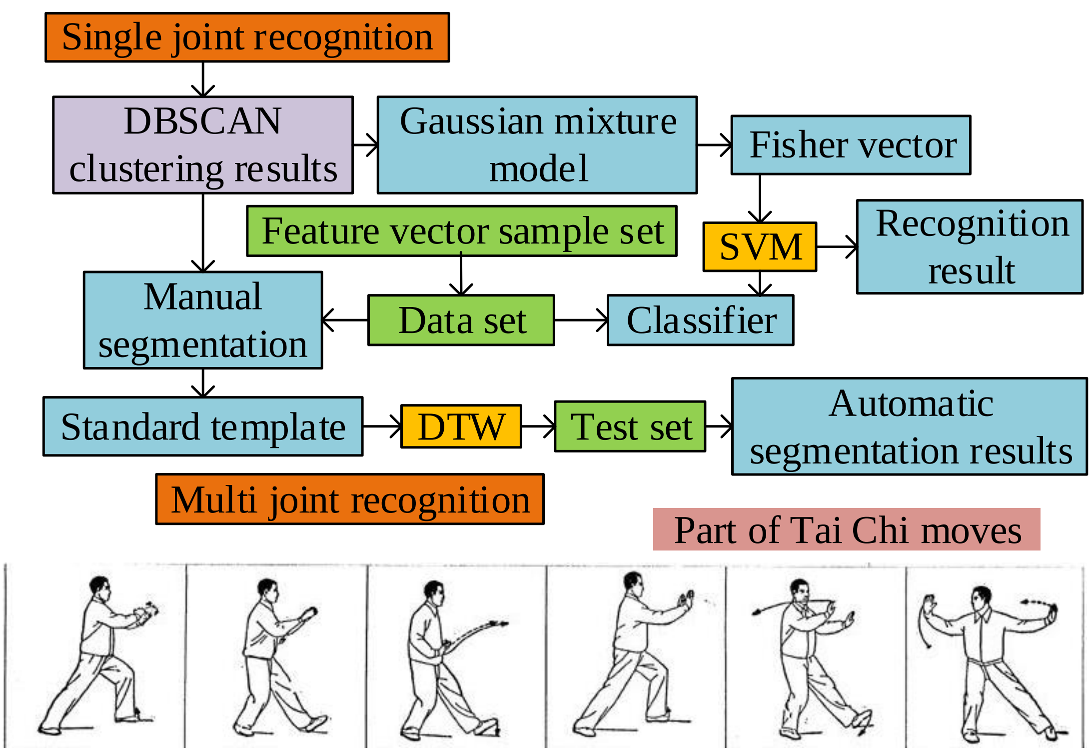


Fig. 5. Schematic diagram of recognition steps for Tai Chi movement trajectory.

IV. Segmentation and Recognition Analysis of Tai Chi Movement Trajectories

Based on the segmentation and trajectory processing of Tai Chi movements, clustering evaluation indicators are used to analyze the clustering effect of the movement trajectories for the left- and right-hand joints in Tai Chi. The step size variation range of the two trajectories is planned to be 0-100. The relationship between the DI evaluation index of hand movement trajectory and the weight sum of parallel, vertical, and angle distances is shown in Fig. 6.

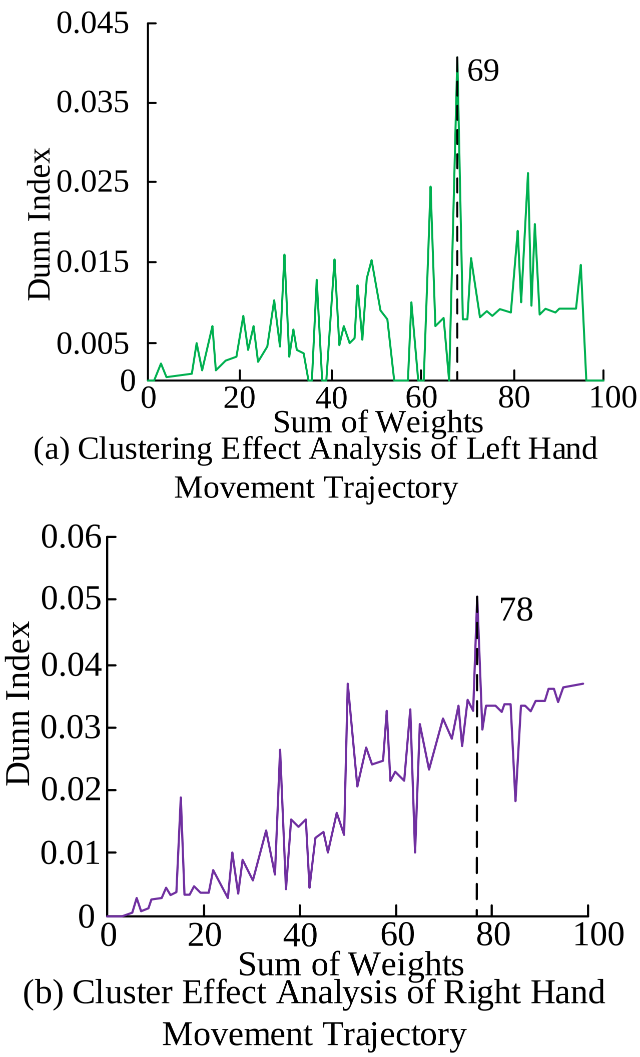


Fig. 6. Clustering effect analysis of hand movement trajectories in Tai Chi.

From Fig. 6 (a), the DI value of the left-hand movement trajectory reached the highest value, indicating the best trajectory clustering effect. The weight sum was 69, and the overall movement trajectory changed differently. When the clustering effect of the right-hand movement trajectory reached its maximum, the weight sum was 78. The overall hand movement trajectory showed an upward trend, that is, the DI value continued to increase overall. The hand movement in Tai Chi is called Yunshou. Due to the clustering processing of the DBSCAN algorithm, it is converted into a Fisher vector, as shown in Fig. 7.

From Fig. 7, the motion trajectory of the Cloud Hand move varied greatly. The Fisher vector changed with increasing size. The overall trend was concentrated between -0.2-0.2, with the highest approaching 0.6. There are many schools of Tai Chi. Fist types and movements are diverse and varied. The study takes the Tai Chi of the Wu school as an example. Ten of the movements, including Grasp the Bird's Tail, Single Whip, Cross Hands, Cloud Hand, Sea Needle, Flash the Arm, Oblique Flying, Upside Down Chasing the Monkey, Tai Chi Qi Shi, and Parry and Punch, are selected for analysis. The names of these movements are represented in order by A-J. Finally, the trajectory of Tai Chi movements is identified and analyzed. The right-hand movement trajectory is recognized as a single joint action. Ten moves are classified into two categories and the training sample proportions are set to 25%, 50%, and 75%, respectively. The recognition accuracy of the two categories is calculated separately. When the training samples account for 25% of the total samples, some recognition results are shown in Fig. 8.

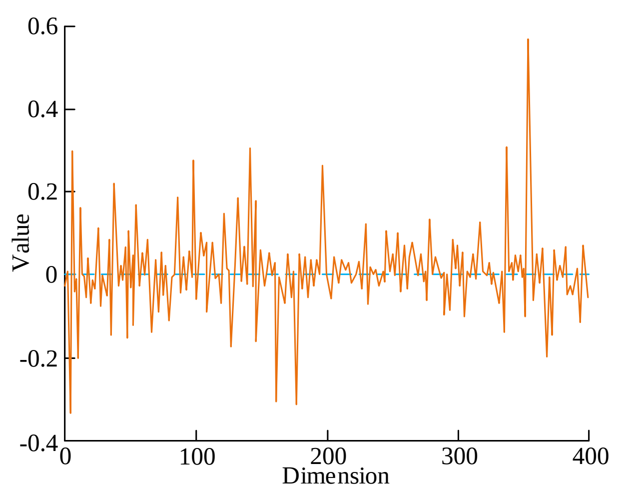


Fig. 7. Fisher vector changes in the trajectory motion of the cloud hand move.

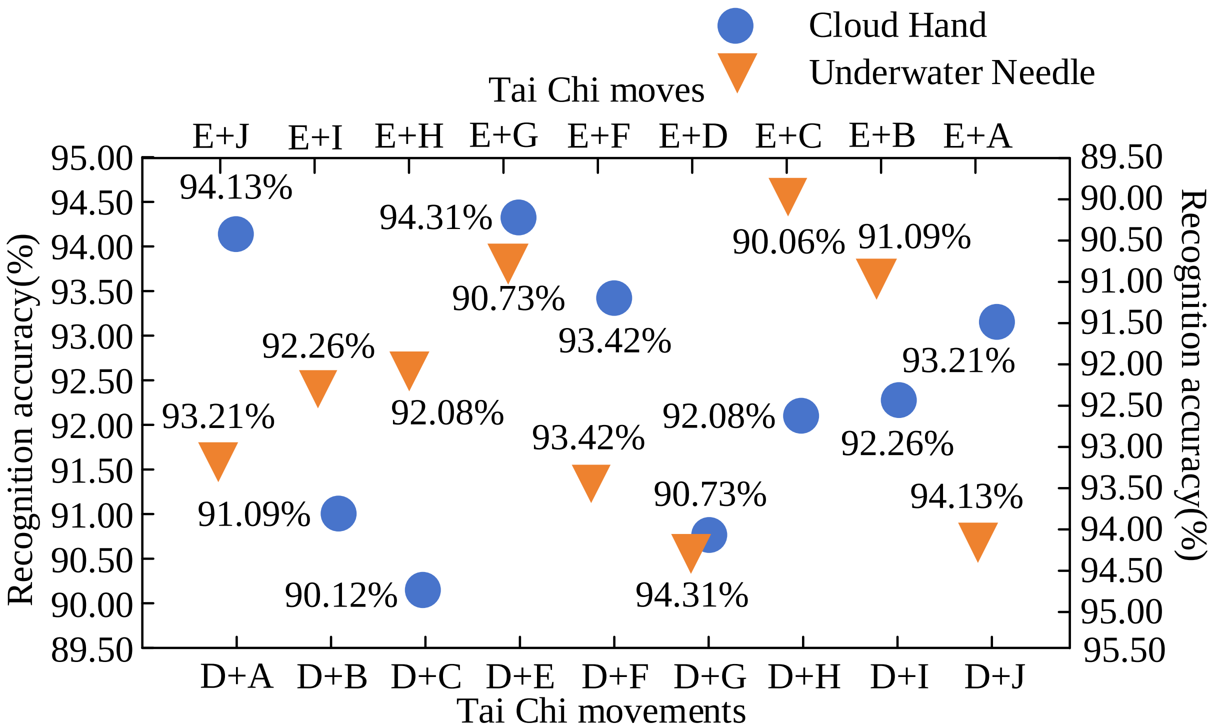


Fig. 8. Recognition results of the 25% binary classification of the cloud hand and sea needle.

From Fig. 8, when the test sample was 25%, the overall recognition accuracy of the Cloud Hand and Sea Needle movements was above 90%, with the highest being 94.31% and the lowest being 90.12% and 90.06%, respectively. When the test sample was 50%, the binary recognition results of some movements are shown in Fig. 9.

From Fig. 9, when the training sample was 50%, the recognition rates of the Flash the Arm, Oblique Flying movements were both above 93%, which was higher than the 25% training sample proportion result. The highest recognition rates for the combination of the Flashing Back and the Oblique Flying were both 98.45%, while the lowest recognition rates were 93.30% and 93.27%, respectively. Finally, the training samples accounted for 75%. The recognition rate changes are tested, as shown in Fig. 10.

From Fig. 10, when the training sample was 50%, the recognition rates of the Flash the Arm, Oblique Flying moves were above 93%, which was higher than the 25% training sample proportion result. The highest recognition rates for the combination of the Flashing Back and the Oblique Flying were 98.45%, while the lowest recognition rates were 93.30% and 93.27%, respectively. Finally, the training samples accounted for 75%. The recognition rate changes are tested, as shown in Fig. 10. This indicates that the sample proportion is large, and the recognition accuracy of the moves continues to improve, thereby verifying the effectiveness of the classification recognition algorithm. To visually compare the binary classification recognition results of different sample proportions, a movement called Qishi is selected to perform 25%, 50%, and 75% binary classification recognition. The change results are shown in Table I.

From Table I, the lowest recognition accuracy for the binary classification of the Qishi movement was 90.87%, 93.53%, and 98.08%, respectively, when the training sample proportions were 25%, 50%, and 75%. Furthermore, it indicates that as the proportion of samples increases, the recognition accuracy also improves. Finally, the binary and multi-classification samples are trained to recognize single joint and multi-joint movements. The accuracy and standard deviation of the average value are taken to determine the classification accuracy of the motion trajectory, as shown in Fig. 11.

From Fig. 11(a), the recognition accuracy of binary classification for single node in the sample proportion of 25%, 50%, and 75% was 92.26%, 96.51%, and 98.48%, respectively. The recognition accuracy in multi-classification was 90.05%, 90.58%, and 93.56%, respectively. In Fig. 11(b), the recognition standard deviations of binary classification for single node in the 25%, 50%, and 75% sample proportions were 1.15%, 1.32%, and 0.47%, respectively. The standard deviations for multi-classification recognition were 2.68%, 2.11%, and 3.08%, respectively. Furthermore, it indicates the variation of single joint points in different sample proportions. The binary classification method has higher recognition accuracy and lower standard deviation than the multi-classification method. This proves that the binary classification method for identifying single related nodes has advantages. The recognition accuracy and standard deviation results for multiple joint points are shown in Fig. 11(c). The recognition accuracy of binary classification and multi-classification was 99.77% and 90.25%, respectively. The identification standard deviations were 0.16% and 0.66%, respectively. This indicates that binary classification method is superior to multi-classification method in multi-joint point recognition methods. This proves the superiority of the SVM-based recognition and classification method.

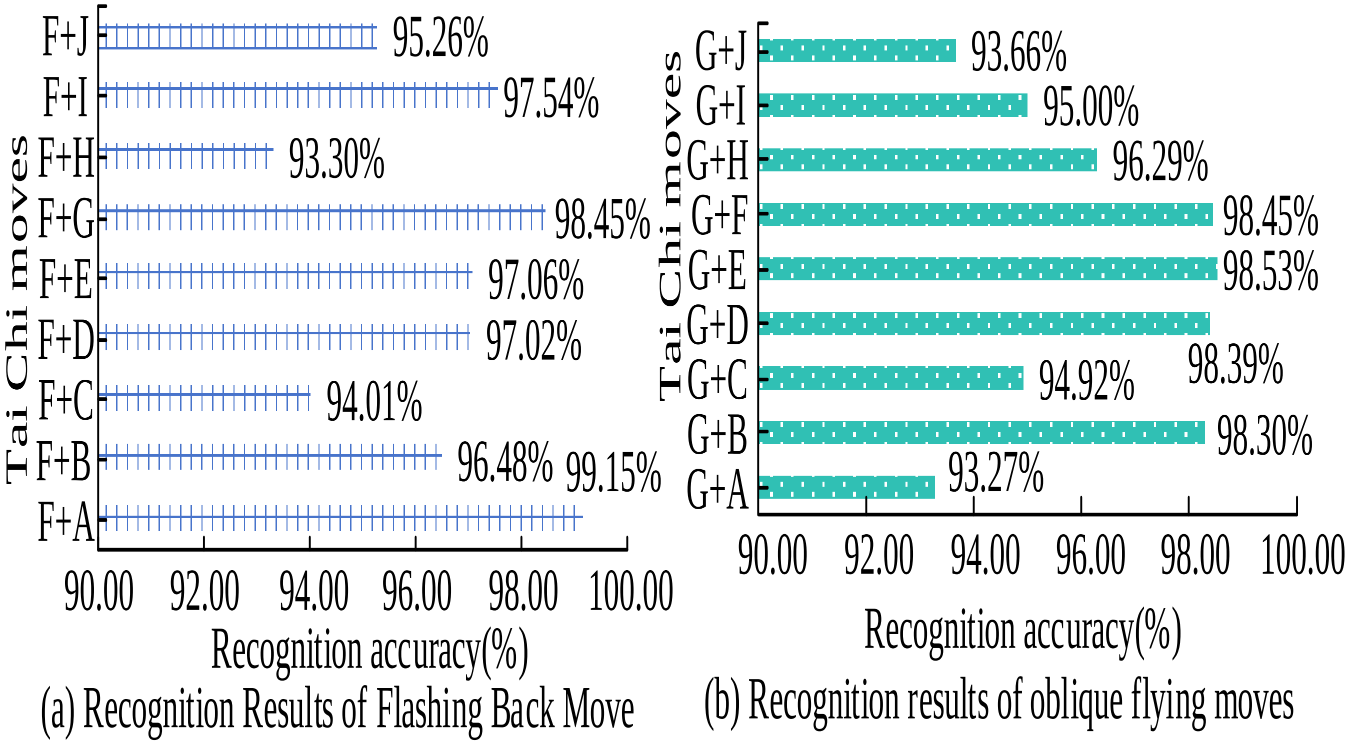


Fig. 9. Recognition results of 50% binary classification for "Flashing Back" and "Oblique Flying".

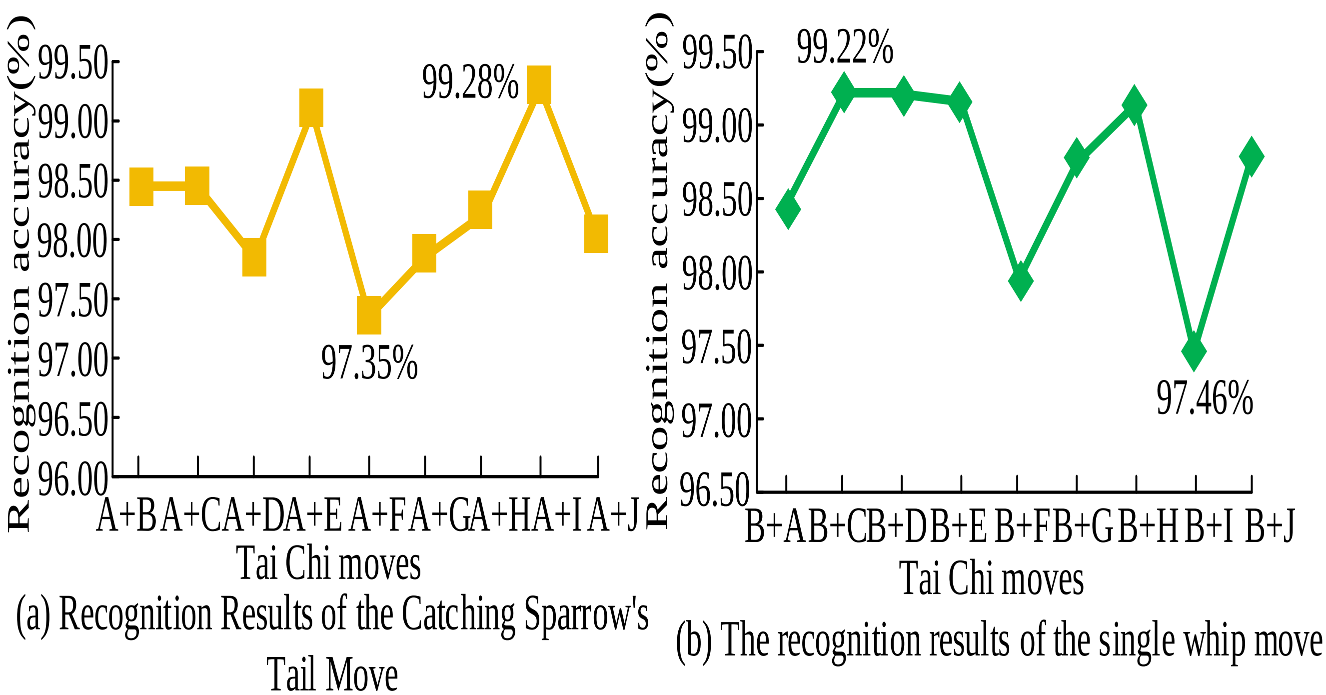
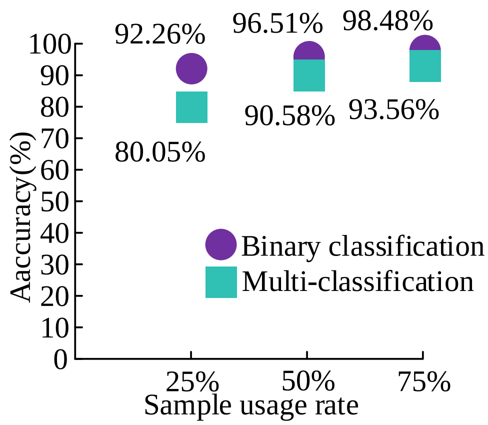


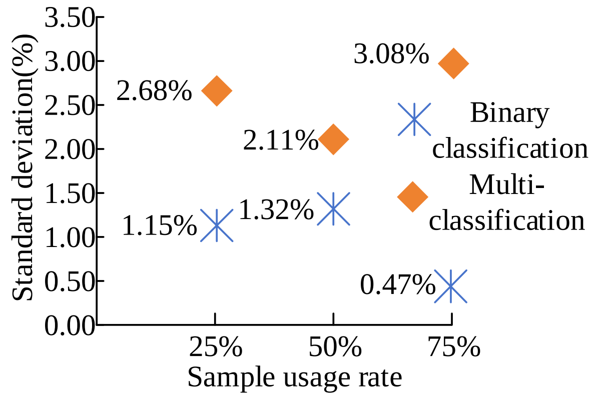
Fig. 10. Recognition results of 75% binary classification for the "Grasping Sparrow Tail" and "Single Whip".

TABLE I. RECOGNITION RESULTS OF THE 25%, 50%, AND 75% DI-CATEGORIZATION OF QI SHI MOVES

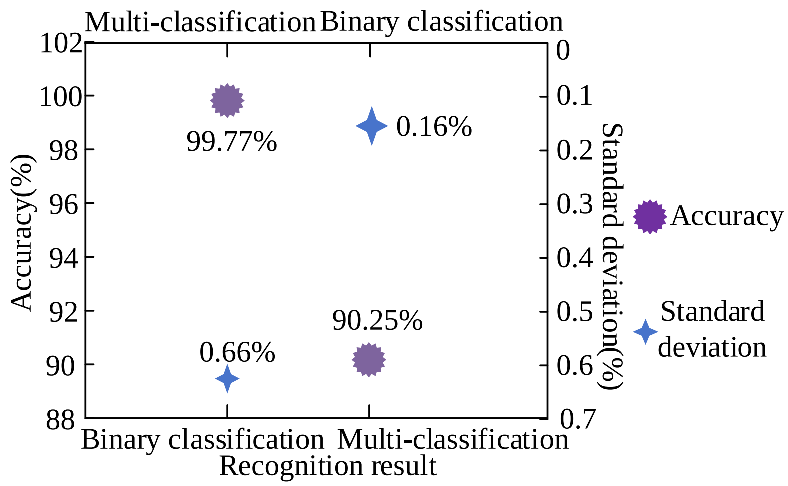
Tai Chi movements	A+I	C+I	D+I	H+I	J+I
25% binary recognition results	94.42%	92.78%	92.20%	92.84%	90.87%
50% binary recognition results	98.02%	96.97%	97.74%	93.53%	98.31%
75% binary recognition results	99.28%	98.08%	98.11%	99.27%	98.55%



(a) Recognition accuracy of single joint binary classification and multi-classification



(b) Recognition standard deviation between single joint binary classification and multi-classification



(c) The recognition accuracy of multi node binary classification and multi-classification

Fig. 11. Recognition results of binary and multi-classification for single node and multi-joint points.

TABLE II. COMPARISON RESULTS OF DIFFERENT METHODS FOR RECOGNIZING TAI CHI MOVEMENTS

Tai Chi movements	E+A	E+B	E+C	E+D	E+F
Reference[6]	90.84%	91.03%	90.53%	89.87%	90.46%
Reference[7]	91.23%	90.45%	90.07%	91.16%	90.32%
Reference[9]	87.61%	86.26%	86.01%	89.26%	88.04%
This research method	99.16%	98.57%	99.12%	98.35%	99.24%

Finally, the research method is compared with existing methods, focusing on trajectory recognition of five Tai Chi moves mainly using the Hai Di Zhen technique. The results are shown in Table II.

From Table II, the deep neural networks and adaptive multi-scale convolutional networks in references [6] and [7] had better recognition methods for human joint points, and had good recognition results for light and gentle Tai Chi movements. The deep network model achieved a recognition result of 91.03% for underwater needles and single whips, while the model in study [7] achieved a recognition result of 91.16% for underwater needles and cloud hands. However, the recognition in study [9] was below 90%, which was not suitable for recognizing Tai Chi movements. Based on the above comparison results, it is concluded that the proposed method has accuracy and applicability in recognizing Tai Chi movements.

V. CONCLUSION

For the analysis of Tai Chi movement recognition data, human motion data is processed by inertial sensors and visual sensors. Then, the collected trajectory data is segmented and clustered. Finally, a comparative analysis is conducted on the recognition of Tai Chi movements. The clustering evaluation index is used to analyze the clustering effect of the movement trajectories for the left- and right-hand joints in Tai Chi. When the clustering effect of left- and right-hand movements was the highest, the sum of weights was 69 and 78, respectively. The specific movements of Tai Chi are used for action recognition. In the binary classification method, when the training sample was 25%, the overall recognition accuracy of the Cloud Hand and Sea Needle moves was above 90%, with the lowest being 90.06%. When the training sample was 50%, the recognition rates of the flashback and oblique flying movements were both above 93%, with the highest being 98.45%. When the training sample was 75%, the recognition accuracy of the "Grasp the Bird's Tail" and "Single Whip" were both above 97%. As the number of training samples increases, the accuracy of action recognition continues to improve. This also indicates the classification effectiveness of the DBSCAN algorithm. Finally, single joint and multi-related nodes were respectively used in binary classification and multi-classification. The recognition accuracy of single joint in the 25%, 50%, and 75% sample proportion of binary classification was 92.26%, 96.51%, and 98.48%, respectively. The recognition accuracy of multi-joint points in binary classification was 99.77%. Furthermore, it proves that the binary classification method has high recognition accuracy, indicating the superiority of the trajectory segmentation and recognition classification method based on multiple sensors. However, human motion trajectory segmentation still lacks a large amount of motion data. In addition, the recognition and selection of Tai Chi movements are not widely applicable. In the action decomposition of sports events, real-time action recognition is not considered in order to collect and correct actions. In future research, it is necessary to reference other types of sports to improve athletes' movement skills and expand the feasibility and practicality of sports movement recognition technology. The accuracy of action recognition in the medical field can help physicians diagnose diseases and provide real-time and effective evaluation of treatment plans for patients' exercise rehabilitation. With the development of intelligence, in the fields of human-computer interaction and virtual reality, the recognition of human movements can be transformed into instruction input, thereby promoting the intelligence of life.

This study analyzed the coherence and overall coordination of Tai Chi movements, and conducted segmentation and clustering. In addition, considering the motion joints of actions, SVM and DTW methods were used to cluster single joints and multiple joints. The DBSCAN algorithm accurately identified the data noise of high-density motion trajectories, thereby improving the clustering effect of trajectories while segmenting them, and ultimately achieving high classification and recognition accuracy of Tai Chi movements.

CONFLICT OF INTEREST

The authors report there are no competing interests to declare.

REFERENCES

[1] Yang M, Wu C, Guo Y, Jiang R, Zhou F, Zhang J, Yang Z. Transformer-based deep learning model and video dataset for unsafe action identification in construction projects. Automation in construction, 2023, 146(2): 104703-104716.

[2] Bhogal R K, Devendran V. Action Recognition for Multiview Skeleton 3D Data Using NTURGB+D Dataset. Computer Systems Science and Engineering, 2023, 47(12):2759-2772.

[3] Haoyu Z, De Z. Combining Adaptive Graph Convolution and Temporal Modeling for Skeleton-Based Action Recognition. Computer Engineering and Applications, 2023, 59(18): 137-144.

[4] Pham D T, Pham Q T, Nguyen T T, Le T L, Vu H. A lightweight graph convolutional network for skeleton-based action recognition. Multimedia tools and applications, 2023, 82(2): 3055-3079.

[5] Sun B, Kong D, Zhang W, Jia W. Survey on Human Action Recognition from Depth Maps. Tsinghua Science and Technology, 2022, 27(6): 29.

[6] Luo F, Khan S, Huang Y, Wu K. Binarized Neural Network for Edge Intelligence of Sensor-Based Human Activity Recognition. IEEE transactions on mobile computing, 2023, 22(3): 1356-1368.

[7] Liu K, Xiaobing X I, Zhou M. Skeleton Action Recognition Based on Adaptive Multi-scale Graph Convolutional Network. Computer Engineering, 2023, 49(10): 264-271.

[8] Liu P, Zhang G, Zhang S, Li Y, Zeng Z. Skeleton-aware implicit function for single-view human reconstruction. Journal of Intelligent Technology, 2023, 8(2): 379-389.

[9] Liu Y. Athletes' Throwing Action Recognition Method Based on PCA-LBP Algorithm. International Journal of Computational Intelligence Studies, 2023, 12(1): 130-141.

[10] Host K, Ivai-Kos M. An overview of Human Action Recognition in sports based on Computer Vision. Heliyon, 2022, 8(6): e09633.

[11] Song Y, Fan C. Behavior Recognition of the Elderly in Indoor Environment Based on Feature Fusion of Wi-Fi Perception and Videos. Journal of Beijing Institute of Technology, 2023, 32(2): 142-155.

[12] Cai X, Xi M, Jia S, Xu X, Wu Y, Sun H. An Automatic Music-Driven Folk Dance Movements Generation Method Based on Sequence-To-Sequence Network. International journal of pattern recognition and artificial intelligence, 2023, 37(5):2358003-2358023.

[13] Chang Y, Wang L, Zhao Y, Liu M, Zhang J. Research on two-class and four-class action recognition based on EEG signals. Mathematical biosciences and engineering: MBE, 2023, 20 6: 10376-10391.

[14] Seo K, Shibata S, Hirose K, Naruo T, Shimizu Y. Estimation of baseball bat trajectory during a practice swing using a Kalman filter for velocity compensation. Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology, 2023, 237(2): 96-101.

[15] Tang Z, Jia C, Wang H, Rong S, Zhao W. Intelligent height measurement technology for ground encroachments in large-scale power transmission corridor based on advanced binocular stereovision algorithms. IET Generation, Transmission & Distribution, 2023, 17(2): 448-460.

[16] Xiao Nan X Y W L. Research on target positioning mode of intelligent unmanned vehicle based on multi-vision sensor. Manufacturing Automation, 2023, 45(3): 76-80.

[17] M. Hasanvand, M. Nooshyar, E. Moharamkhani, and A. Selyari. "Machine Learning Methodology for Identifying Vehicles Using Image Processing," AIA, 2023, 3(1): 170-178.

[18] Naji O A A M, Shah H N M, Anwar N S N, Johan N F. Square groove detection based on Frstner with Canny edge operator using laser vision sensor. The International Journal of Advanced Manufacturing Technology, 2023, 125(5-6): 2885-2894.

[19] Xi Z. Research on the video motion segmentation and its application. Acta Geodaetica et Cartographica Sinica, 2023, 52(8): 1411-1411.

[20] Xu Y, Chen T, Chen S, Xu X. Multi-object Tracking and Segmentation Algorithm by Fusing Motion Feature Embedding. Journal of Chinese Computer Systems, 2023, 44(6): 1304-1310.