



Article

Using Adaptive Directed Acyclic Graph for Human In-Hand Motion Identification with Hybrid Surface Electromyography and Kinect

Yaxu Xue ¹ , Yadong Yu ¹, Kaiyang Yin ¹ , Haojie Du ^{1,*}, Pengfei Li ¹, Kejie Dai ¹ and Zhaojie Ju ²

¹ School of Electrical and Mechanical Engineering, Pingdingshan University, Pingdingshan 467000, China

² School of Computing, University of Portsmouth, Portsmouth PO1 3HE, UK

* Correspondence: 2629@pdsu.edu.cn

Abstract: The multi-fingered dexterous robotic hand is increasingly used to achieve more complex and sophisticated human-like manipulation tasks on various occasions. This paper proposes a hybrid Surface Electromyography (SEMG) and Kinect-based human in-hand motion (HIM) capture system architecture for recognizing complex motions of the humans by observing the state information between an object and the human hand, then transferring the manipulation skills into bionic multi-fingered robotic hand realizing dexterous in-hand manipulation. First, an Adaptive Directed Acyclic Graph (ADAG) algorithm for recognizing HIMs is proposed and optimized based on the comparison of multi-class support vector machines; second, ten representative complex in-hand motions are demonstrated by ten subjects, and SEMG and Kinect signals are obtained based on a multi-modal data acquisition platform; then, combined with the proposed algorithm framework, a series of data preprocessing algorithms are realized. There is statistical symmetry in similar types of SEMG signals and images, and asymmetry in different types of SEMG signals and images. A detailed analysis and an in-depth discussion are given from the results of the ADAG recognizing HIMs, motion recognition rates of different perceptrons, motion recognition rates of different subjects, motion recognition rates of different multi-class SVM methods, and motion recognition rates of different machine learning methods. The results of this experiment confirm the feasibility of the proposed method, with a recognition rate of 95.10%.

Keywords: multi-fingered hand; human in-hand manipulation (HIM); surface electromyography (SEMG); Kinect; adaptive directed acyclic graph (ADAG)



Citation: Xue, Y.; Yu, Y.; Yin, K.; Du, H.; Li, P.; Dai, K.; Ju, Z. Using Adaptive Directed Acyclic Graph for Human In-Hand Motion Identification with Hybrid Surface Electromyography and Kinect. *Symmetry* **2022**, *14*, 2093. <https://doi.org/10.3390/sym14102093>

Academic Editors: Deming Lei and Ming Li

Received: 9 August 2022

Accepted: 1 October 2022

Published: 8 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As an intelligent machine that performs various tasks in place of human beings, robots play an increasingly significant role in the production and life practices of human beings. Various types of intelligent robots, such as surgical robots, inspection robots, and welding robots, have been designed based on specific task requirements [1]. As the end effector for various operations, the manipulator mainly includes three parts: hand, motion mechanism, and control system. According to the physical properties of operated objects, including color, size, weight, and texture, as well as the operating requirements, there are three major types of manipulators: mechanical gripper, adsorption end-effector, and specialized tools (brazing torch, nozzle, electric grinding head, etc.). Because of superior persistence and a certain degree of dexterity, manipulators can not only improve working conditions, but can also enhance the labor productivity and product quality of large-scale production, thus bringing enormous economic benefits. However, conventional manipulators have significant limitations when faced with complex operating environments and requirements, and the main drawbacks are listed as follows:

- There are lower gripping accuracy, poorer stability, and reliability in the absence of geometric closure and force closure;

- It is difficult to achieve precise orientation and operation, as well as the poor dynamic response;
- It can't fulfill the tasks that require high grasping force with a lack of precise force control.

By observing and studying the dexterity of human hands, more and more researchers have developed a strong interest in designing multi-fingered robotic hands that are similar in structure and function to human hands for realizing complex manipulation and control in various application environments [2,3]. Hence, how to transfer HIM skills to a multi-fingered robotic hand is a research hotspot in the field of artificial intelligence (AI) nowadays. The human hand, as the most flexible and complex structure of the human body, can finish about 90 percent of the operational tasks in human daily life. Humans can make the corresponding reasonable motions performed with the hand and fingers according to the different tasks. In addition, hand motions are also affected by the fact that different operators have different manipulations for the same task. More and more researchers have been attracted to hand manipulation skill analysis and, further, to apply them to control bionic multi-fingered dexterous robotic hands [4–7]. Elliott et al. first proposed a manipulation skills classification framework to describe four broad classes of hand motions in detail [8]. Based on Elliott's research on the classification of human hand motions, Exner divided human hand movements into five categories [9]. Pont et al. further summarized and analyzed previous human hand action classification methods and proposed a new human hand action classification method including six classification modes [10].

More recently, Xue et al. proposed a generalized framework using multiple sensors to study and analyze ten types of commonly used hand manipulations [11]. Bullock et al. presented a hand motion manipulation classification method to create a descriptive framework for classifying patterns of different dexterous manipulations involved in each task [12]. From the description of human hand manipulation in the above-mentioned related literature, it can be seen that the object operation is centered on the human hand, and a series of operation sub-actions are generated through the explicit interaction between the human hand and the object [13]. Before humans grasp an object reasonably, they first select the number of fingers and grasping model reasonably according to the specific requirements of the operation task and the physical characteristics of the object, and then select the best contact point based on past experience to achieve stable manipulation [14]. However, the arm and hand joints provide the necessary position information and force information for the stable operation of objects. Therefore, based on the summary of the above human hand manipulation modes and the analysis of human hand actions and object interactions, this paper designs ten typical human hand motions to represent the feature information included in different actions.

Considering that human hand manipulation is a dynamic process that includes continuous sub-actions and is affected by personal habits, posture changes, etc., resulting in different human actions, it will be a huge challenge to obtain the operation information of human hand motions. Researchers have begun to analyze the dynamic process of human hand manipulation from a biological perspective. Generally speaking, the brain first generates an electrical signal of human hand motion intention, and the nervous system transmits an electrical signal and stimulates the corresponding muscle fibers. Then, a complex set of skeletal muscles, tendons, and bones are used to bring forth various hand motions. A variety of sensors are currently used to capture the physical characteristics of human hand motions, involving finger force, angle, position, and other information. Surface electromyography (SEMG) can accurately collect the SEMG signals generated by muscle fiber contraction and is widely used in human hand motion mechanics analysis, gait analysis, clinical rehabilitation, and other fields. Therefore, a large number of research articles show the research results of SEMG signal processing [15–17]. Because of the nonlinear and chaotic behavior of SEMG signals [18], nonlinear time series analysis, including the Markov Model (MM), Threshold Autoregressive Model (TAM), Lyapunov Spectrum (LS), and other proposed algorithms, are the classical and popular methods used for feature extraction. In order to better capture the spatial and image information of HIMs, the most widely used

visual perception device with perfect image information is the Kinect sensor developed by Microsoft nowadays [19]. It can acquire the depth image information of human hand motions in unstructured environments, avoiding the error interference caused by physical contact. Ju et al. designed a Kinect-based human hand gesture recognition framework for ten types of nuanced human hand motion classification using the Finger-earth Movers Distance (FMD) method [20]. Using the 2G-Kinect sensor, Sun et al. designed a 3D skeleton gait data set, including 2D contour images of human hand movements and 3D coordinates of skeleton joints [21]. In addition, some representative articles on vision-based human hand motion recognition and technology applications have been published in a wide range of international robotics journals and conferences.

Considering the complexity of human hand movements, a single sensor cannot acquire multiple physical features; therefore, the lack of key features will lead to motion distortion of HIM recognition. It is necessary to adopt the multi-modal data fusion method to make up for the disadvantage that a single perceptron can only obtain uni-modal data. By reviewing the state-of-the-art literature, there are few researchers have simultaneously acquired SEMG signals and 3D depth information of human hands based on SEMG and Kinect sensors. This paper proposes an HIM recognition framework based on two different kinds of sensory information, which include a SEMG signal from a high SEMG capture system attached on the forearm, and depth-sensing information from Kinect placed on the physical desktop. In addition, a novel threshold-based human hand motion segmentation algorithm is proposed; that is, the threshold value is calculated by sampling the maximum and average absolute values of the SEMG signal to judge the starting point of different human hand actions and realize motion segmentation.

The remainder of the paper is organized as follows. First, the detailed HIM identification method is presented in Section 2, and the HIM capture system architecture, including system principles, acquisition platforms, and motion capturing, is proposed in Section 3. Then, the proposed multi-modal signal processing method is introduced in Section 4. Section 5 mainly presents the experimental results analysis and discussion with different comparison methods. The final Section summarizes the proposed HIM recognition method and further research direction.

2. Human In-Hand Motion Identification Method

2.1. Multiclass Support Vector Machines

HIM recognition belongs to the category of multi-classification. It refers to obtaining the original human hand behavior feature information and then obtaining a feature set with better representation ability and separation degree through multimodal data analysis. The common multiclass support vector machines (MSVM) are mainly divided into four categories, as shown in Table 1 [22,23]. By comparing the merits and demerits of the four methods, this paper propounds a novel MSVM method, namely adaptive directed acyclic graph (ADAG), which can effectively avoid misclassification and rejection, and achieve high-precision and accurate classification.

Table 1. Comparison of four MSVM methods.

Methods	Training Time	SVs	Size	Merits	Demerits
One-versus-rest	Short	K	Large	Simple Effective	Misclassification
One-versus-one	Long	$K*(k - 1)/2$	Moderate	High accuracy	Inseparable problem
Direct MSVM	Long	No SVs	Large	Natural optimization	Complex computation
DAGSVM	Long	$K*(k - 1)/2$	Large	Efficient to train Avoid misclassification No rejected classification	Error accumulation

2.2. Decision DAGSVM

It can be seen from Table 1 that although one-versus-rest and one-versus-one methods have different training times, SVs, and sample sizes, both of them all have the problem of inseparable regions. Assume that the linear classification model of the binary class is

$$Y_{ij}(x) = w_{ij}^T \varphi(x) + b_{ij} \quad (1)$$

where w_{ij} and $\varphi(x)$ represent a y -dimensional vector and a mapping function that maps x into y -dimensional feature space respectively. Furthermore, b_{ij} means a bias term, and $Y_{ij}(x) = -Y_{ji}(x)$. The inseparable region (brown part) is shown in Figure 1 with three classes, and its pairwise formulation is given by:

$$R_i = \left\{ x \mid Y_{ij}(x) > 0, j = 1, 2, \dots, n, j \neq i \right\} \quad (2)$$

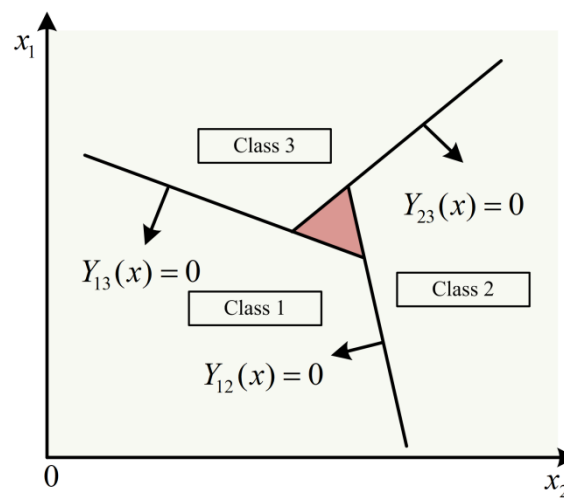


Figure 1. Inseparable region.

Suppose that any sample point x belongs to R_i , then x will be classified into category x -th, otherwise it will be classified into the category with the most votes. The formula of the classification idea is:

$$Y_i(x) = \sum_{j=1, j \neq i}^n \text{sign}(Y_{ij}(x)) \quad (3)$$

Here, we use the symbolic function to determine the range of the input vector x , and the formula of the symbolic function is

$$\text{sign}(x) = \begin{cases} 1 & x \geq 0 \\ -1 & x < 0 \end{cases} \quad (4)$$

It is worth noting that when $x = 0$, $\text{sign}(x) = 1$. Through Formulas (3) and (4), the classification expression of any sampling point x is

$$\arg \max_{i=1, \dots, n} Y_i(x) \quad (5)$$

Suppose $x \in R_i$, $k \neq i$, $Y_i(x) = n - 1$ and $Y_k(x) < n - 1$, then x will be classified into i . It should be noted that if all $Y_i(x) \neq n - 1$, plural is will appear in Formula (5), and x is unclassifiable. Therefore, the brown part in Figure 1 is an indivisible area, that is $Y_i(x) = 0 (i = 1, 2, 3)$.

The decision DAG is only one direction from top to bottom in the graph, and no cycles appear. All nodes are arranged and distributed in an equilateral triangle; that is, the number of nodes in the first layer is 1, the number of nodes in the second layer is 2, and so

on. DAG-SVM starts from the root node and judges the classification of the sample points through the two-parameter function on the node until only one class is included in the list. Hence, the inseparability, misclassification, and rejection of multi-classification problems are resolved, as shown in Figure 2.

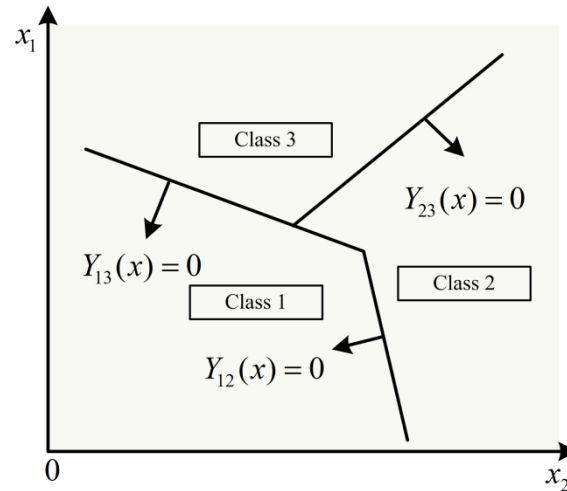


Figure 2. Directed acyclic graph classification.

2.3. DAGSVM Optimization

The DAGSVM method constructs the k -class SVM classification problem into k two-class classifiers. The i -th SVM classification problem treats the i -th class as one class and the rest as another class, thereby transforming the multi-classification problem into a local binary classification problem. For a sample $(x_1, y_1), \dots, (x_n, y_n)$ with n training data, where $x_i \in R^D, y_i \in \{1, \dots, k\}$ is the class label of $x_i, i = 1, 2, \dots, n$. The i -th SVM needs to solve the following optimization problem:

$$\min_{w, b^i, \varepsilon^i} \frac{1}{2} (w^i)^T w^i + C \sum_{j=1}^n \varepsilon_j^i (w^i)^T \quad (6)$$

$$\begin{cases} (w^i)^T \phi(x_i) + b^i \geq 1 - \varepsilon_j^i, & y_j = i \\ (w^i)^T \phi(x_i) + b^i \leq \varepsilon_j^i - 1, & \text{else} \end{cases} \quad \varepsilon_j^i \geq 0, \quad j = 1, \dots, n \quad (7)$$

The decision function obtained from Formula (7) is:

$$(w^1)^T \phi(x) + b^1, \dots, (w^k)^T \phi(x) + b^k \quad (8)$$

suppose x belongs to the category with the largest output value of the decision function, and its expression is

$$x = \arg \max_{i=1, \dots, k} \left((w^i)^T \phi(x) + b^i \right) \quad (9)$$

When the training sample classes are large, the training samples of a certain class are much less than the sum of the training samples of other classes. Due to the imbalance of the training samples, there will be misclassification and rejection areas, which will affect the final classification accuracy.

In order to solve the misclassification and rejection area problems of DAGSVM, this paper proposes an Adaptive Directed Acyclic Graph (ADAGSVM) method to reduce the dependence on the node order and the hierarchical depth of the directed acyclic graph, so as to achieve correct classification of sample data. The ADAG method adopts a hierarchical decision tree structure with an inverted triangular structure, and the training method is the same as that of the decision-directed acyclic graph. When the ADAGSVM algorithm

deals with a k classification problem, it needs $k * (k - 1)/2$ binary classifier and $k - 1$ internal nodes. Figure 3 is a schematic diagram of the ADAG algorithm processing an 8-classification.

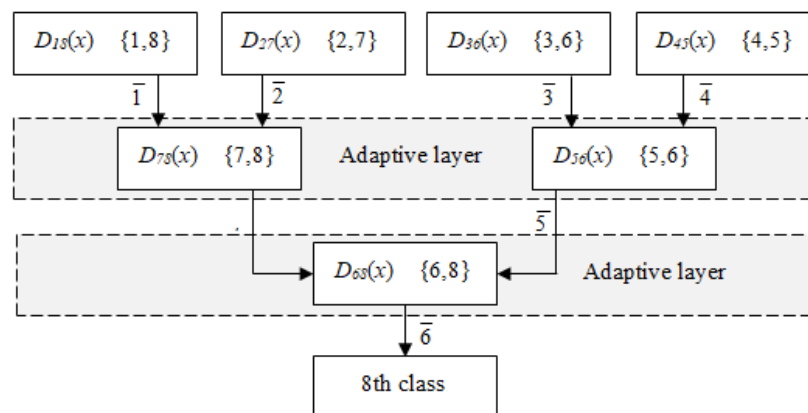


Figure 3. The classification principle of ADAGSVM.

In Figure 3, the first layer contains 4 nodes and evaluates the two-parameter function of each node from this layer and identifies the class of the samples in each node through the output value of the two-parameter function; after each round, the sample The classes are halved and the two-argument function for the next level node is selected based on the parent's preferred class; the classification process goes all the way down to the lowest level node.

2.4. Model Optimization

When solving a linear inseparable problem, the sample values are mapped to a higher-dimensional space or an infinite-dimensional space through a mapping function. If the method of classification problem in the above linear separable case is used for a new linearly separable sample, it is necessary to calculate the inner product of the samples. Through a comparison of kernel functions and the characteristics of the ADAGSVM model, the Gaussian kernel function is the optimal choice for model optimization in this paper. Suppose χ and E are the input space (Euclidean space or discrete set) and the feature space (Hilbert space), respectively, and the mapping of χ to E is:

$$H(x) = \chi \rightarrow E \quad (10)$$

where $\forall x, y \in \chi$, if $Z(x, y) = \varphi(x) \cdot \varphi(y)$, then $Z(x, y)$ is called the kernel function. $\varphi(x) \cdot \varphi(y)$ is the inner product of (x, y) mapped to E . However, in practical applications, since the mapping function is complex and the sample dimension is very high, it is easy to cause a "dimension disaster", so a kernel function is introduced here to convert the inner product of high-dimensional vectors into low-dimensional vectors.

The penalty parameter C is a compromise between obtaining a wide boundary spacing and a small number of edge singularities for constraining the Lagrangian multipliers. The experiments of a large number of researchers have proved that the method to find the best penalty parameter C is still to try to verify it in an exponential way in one of the following areas:

$$C = 2^{-5}, 2^{-4}, \dots, 2^5 \quad (11)$$

To obtain the ideal core radius γ , it is necessary to formulate reasonable model evaluation criteria. The common cross-validation method is leave-one-out. The main idea of the leave-one-out classification method is to treat a certain category sample in multiple categories as a category, and the remaining samples as another category. The results obtained by this method are relatively reliable, but the number of models to be established is equal to the number of samples, which increases the calculation cost.

This paper proposes a novel grid search method to acquire the best parameters (C, γ) . Since there will be different (C, γ) corresponding to the highest precision, the group of (C, γ) with the smallest penalty parameter is considered as the optimal choice, and then within the range of the group of parameters, through the step-by-step method to find a more ideal parameters (C, γ) . For example, the optimal (C, γ) of motion 2 is (1,0.0237). After obtaining the optimal penalty parameters, the whole training set needs to be trained again to obtain the final classifier. In addition, in order to avoid the problem of overfitting, we used the five-fold cross-validation method to test the data set. From the test results, the average accuracy obtained is good.

3. In-Hand Manipulation Capture System Architecture

3.1. System Principle

This paper proposes a hybrid SEMG- and Kinect-based HIM recognition system for robotic human-like manipulation, as shown in Figure 4. The system includes three parts: motion capturing, multi-modal data processing, and motion recognition. First, the multi-modal data acquisition platform obtains 10 kinds of complex human motion information; then, the sample features are acquired through image segmentation, feature extraction, and other methods; finally, the different complex HIMs are recognized based on the proposed ADAG algorithm.

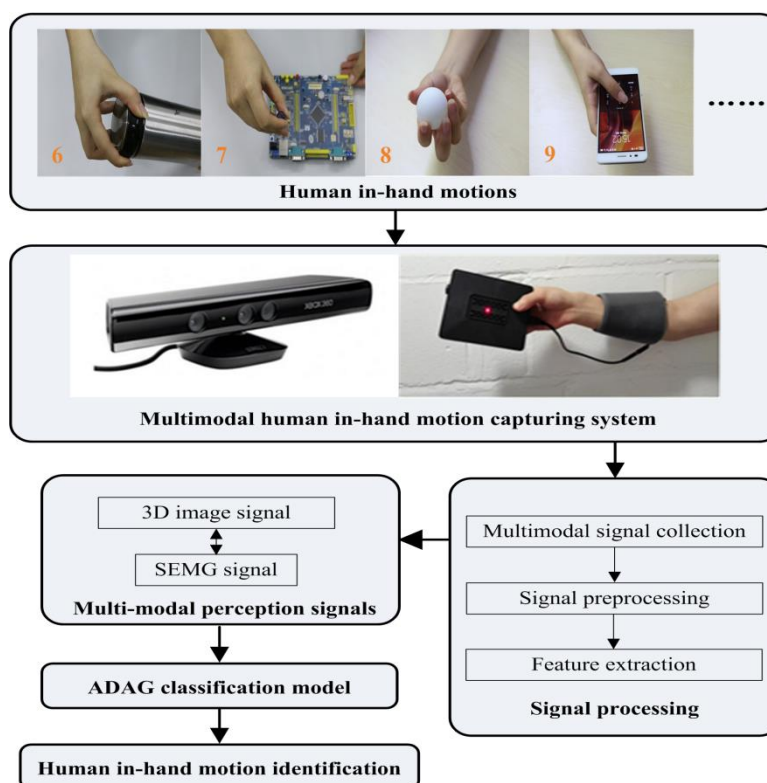


Figure 4. Framework of the HIM recognition system.

3.2. Multi-Modal Data Acquisition Platform

The human hand motion capture platform includes SEMG and Kinect, which can effectively obtain raw data and a sample feature database for identification by tracking human movements. First, according to the complex human hand movements defined by the predefined module of complex human hand movements, the time domain signal acquisition module obtains the sample data F_{SEMG} of the time domain signals and stores the complex human movement feature samples in the feature sample data F . The 3D image information acquisition module obtains the sample data F_{Kinect} of 3D image information and stores it in the feature sample data F of complex human hand movements.

3.2.1. Surface Electromyography

The electromyography instrument uses a surface electromyography signal capture system and Trigno wireless sensor. EMG signal capture of human hand movements is realized through 16 EMG sensors and 48 accelerometers. The external dimensions of the SEMG are 37 mm × 26 mm × 15 mm; the signal resolution and sampling rate are 16 bits and 4000 Hz, respectively; the maximum transmission distance is 40 m; and the maximum charging capacity can maintain continuous operation for 7 h. Each perception unit has a built-in three-axis acceleration sensor for judging the normal force direction; 4 silver bar contacts are applied for SEMG signal capture. In order to correctly use the SEMG device, the arrow mark on top of the device should be parallel to the rotation direction of the muscle fibers below. The sensor should be correctly placed in the center of the muscle abdomen and away from the tendon and its edge to reduce noise interference, such as crosstalk and redundancy.

3.2.2. Kinect Sensor

Kinect is mainly composed of three parts, namely RGB color camera, infrared transmitter, and infrared CMOS camera. The image acquisition frequency of Kinect is set to 30 frames per second, with the red center point as the coordinate center, and a square area of 200 × 200 pixels is saved as a sample image. 1000 sequences of SEMG EMG signals and 2000 color and depth images of Kinect were collected as a sample database.

Kinect can acquire 3 color components (red, green, and blue) of the object simultaneously by 3 different lines and acquire the color signal of the object through image superimposition and color change captured by an independent CCD sensor, in addition to acquiring the object synchronously depth image. The effective detection range of Kinect is about 0.5 m ~ 8 m. Through the complete Kinect for Windows SDK 1.8 development tools provided by Microsoft, you can easily use C++ and other programming languages to control the Kinect and realize the collection of human gestures.

3.3. Motion Capturing

In order to ensure the authenticity and objectivity of complex HIM manipulation, 10 healthy adults were invited as subjects to capture human hand motions. The subjects included 8 men and 2 women, with an average age of 23.7 years, an average height of 171.5 cm, and an average weight of 64.9 kg. None of them had any history of neuromuscular disease, and their right hands were healthy. All subjects received strict operation training before the action capture, mainly including the selection of correct gestures, the location of contact points, and the duration of actions. The subjects had to complete 10 movements, each of which was repeated 10 times, with 5 s of relaxation time between each motion. All subjects signed the Informed Consent Form and obtained ethical approval from the Academic Committee. In this paper, 10 representative HIMs were designed, and the correct manipulation demonstration is shown in Figure 5.

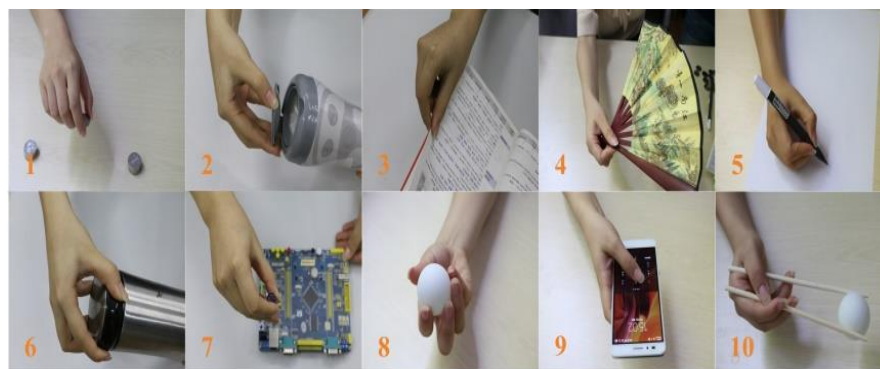


Figure 5. Demonstration of ten customized HIMs.

4. Multi-Modal Signal Processing Method

4.1. Motion Segmentation

How to find the starting and end points of each motion from the signal stream is the main goal of segmentation. The SEMG signals will be segmented synchronously with the Kinect data. There are two states of SEMG signals: transient and steady. The muscle goes from a state of nature to a voluntary contraction level in the beginning. In this transient state, there will be obvious signal fluctuation and error. Hence, the steady-state signal is applied to acquire the best result of data segmentation.

This paper proposes a threshold-based motion segmentation algorithm that obtains the threshold of action segmentation by sampling the maximum and average values of SEMG signals. The definition of threshold T is:

$$T = \begin{cases} \frac{3}{L} \sum_{i=1}^L |x_i| & \max_i \{x_i\} > \frac{30}{L} \sum_{i=1}^L |x_i| \\ \max_i \{x_i\} / 3 & \text{else} \end{cases} \quad (12)$$

where x_i and L represent the input value and length of the SEMG signal, respectively, and the sampling time is 3 s. By sampling the SEMG signal, the allowable range of the threshold T is between 30 ~ 100 μV . When $T = 30 \mu V$, it can meet the requirements of motion segmentation.

When the SEMG device collects the HIM data, the original information after entering the steady state can be regarded as the real effective motion signal, which is defined as:

$$F_p(t, l) = \frac{1}{l} \sum_{i=t-l+1}^t \left[\sum_{j=1}^{16} f_j(i) \right]^2 \quad (13)$$

where $f_j(i)$ (i is the sampling start time) is the i -th sampling value of the j -th channel of the selected EMG signal (j is the number of sampling channels, 1, 2, ..., 16), l is the sampling length. If $x_i \in [0, L]$, the wave criterion is:

$$\begin{cases} (\max\{x_i\} - x_o) \leq T \\ x_i \geq x_o, \quad i \in [o - s, o + s] \end{cases} \quad (14)$$

where x_o is the position of the peak point, and the threshold T is

$$T = (\max\{x_i\} - \min\{x_i\}) \times 0.5 \quad (15)$$

An active segment starts at the p -th point if $F_p(p + s, l)$ at its s -th consecutive point, is larger than T . The values of l and s in this experiment are 200 and 50, respectively, which effectively ensure the motion segmentation of SEMG signals. Kinect will continuously track the HIMs during the collection of useful SEMG signals. 3D scene information is selected from continuously projected infrared structured light as Kinect data for the selected HIMs.

4.2. Feature Extraction

Feature weighting or feature selection plays a key role in extracting useful information from HIM data, including SEMG signals and 3D scene information. It is a set of vectors obtained by reducing the dimensions of the original data. The eigenvector does not affect the representation of the original data and can effectively reduce the amount of calculation, thus reducing the computing time. Next, a detailed presentation of feature extraction from the SEMG signal and Kinect data will be introduced.

4.2.1. Feature Extraction from the SEMG Signal

Selecting representative signal features can more effectively implement pattern classification. Dennis Tkach et al. summarized and defined 11 time-domain feature expressions that can effectively represent the original data and studied their effects and stability during

SEMG signal changes [24]. Time frequency features are widely used to characterize changes in muscle contractility due to their advantages of low computational complexity and strong stability. The feature vector of each SEMG signal we selected includes six types of single feature: mean absolute value (MAV), waveform length (WL), root mean square (RMS), average amplitude change (AAC), zero crossing (ZC) and slop sign change (SSC). The related mathematical equations are presented in Table 2, and the extracted features are collected into F_{SEMG} .

Table 2. Mathematical equations of six features.

Classified Features	Equation
MAV	$\frac{1}{N} \sum_{i=1}^N x_i $
WL	$\sum_{i=1}^{N-1} x_{i+1} - x_i $
RMS	$\sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$
AAC	$\frac{1}{N-1} \sum_{i=1}^{N-1} x_{i+1} - x_i $
ZC	$\sum_{i=1}^N [f(x_i \times x_{i+1}) \cap x_{i+1} - x_i \geq \varepsilon]$
SSC	$\sum_{i=2}^{N-1} f[(x_i - x_{i+1}) \times (x_i - x_{i-1})]$
N : the length of the segment x_i : the i -th sample ε : a threshold	
	$f(x) = \begin{cases} 1, & \text{if } x \geq \varepsilon \\ 0, & \text{otherwise} \end{cases}$

4.2.2. Feature Extraction from the Kinect Signal

RGB images are closely related to brightness, and when the brightness changes, the three color components will change accordingly, and they are not independent of each other. Therefore, the Ycbcr color space is used to realize the image conversion of the binary color space, and the smoothing process is carried out so that the segmentation of the hand area and the background is more obvious, and the edge is smoother. In the Ycbcr color space, Y, cb, and cr represent the luminance component, blue chrominance component, and red chrominance component, respectively. Different from the relationship between RGB color components, the luminance information and chrominance information in Ycbcr are independent of each other, and the chrominance information is not affected by luminance, which has stronger stability and independence, and is more suitable for image segmentation. Therefore, choose the Ycbcr color space as a hand action region color feature. The conversion formula for changing the color space from RGB to Ycbcr is:

$$\begin{cases} Y = 0.257 \times R + 0.564 \times G + 0.098 \times B + 16 \\ cb = -0.148 \times R - 0.291 \times G + 0.439 \times B + 128 \\ cr = 0.439 \times R - 0.368 \times G - 0.071 \times B + 128 \end{cases} \quad (16)$$

Considering the characteristics of the human hand, only the cr component is used as an aid here. Based on the skin color of the yellow race, after consulting data and multiple test results, it is found that the effect obtained is better when the cr component is between 125 and 188. Compared with simple human hand motions, the manipulation requirements of complex HIMs are higher, and multimodal information is more difficult to obtain. The positions and angles of the involved fingers and palms change continuously, and the acquired images of the hand region cannot be directly extracted for features. For this purpose, palm region segmentation needs to be performed from it. palm region $M(s, t)$ is:

$$S = \{X_{s,t} | P(s, t) < V + T\} \quad (17)$$

$$M(s, t) = \begin{cases} 1 & \text{if } X_{s,t} \in F \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

where S is the hand sample set; $X_{s,t}$ is the 3D point obtained by the back-projection of the depth sample at the position (s, t) ; V is the minimum depth value on the threshold depth

image; F is the sample feature set; T is the empirical threshold of the distance between the hand position and the Kinect ($T = 50$ cm is taken in this paper).

Using the feature that the palm area $M(s, t)$ has the highest point density, find a suitable starting point G for the circle fitting to obtain the center point C of the palm area. This paper sets $G = G_g$, G_g is the point with the largest gray value on $M(s, t)$, and there may be multiple G_g . To find a good starting point G , threshold $M(s, t)$ with a binary template $M^T(s, t)$, and its calculation formula is:

$$M^T(s, t) = \begin{cases} 1 & \text{if } M(s, t) \geq T_v \times m_{\max} \\ 0 & \text{otherwise} \end{cases} \quad (19)$$

where $m_{\max} = \max_{s,t}(M(s, t))$ is the maximum calculated density, and $T_v \in [0, 1]$ is the corresponding threshold ($T_v = 0.95$ is taken in this experiment, for example, $T_v \times m_{\max}$ corresponds to 95% of the maximum density). Select the reference point closest to the starting point of the hand detection process as the starting point G_{start} of the circle fitting algorithm.

According to the characteristics of the complex movements of human hands, the image features of the complex movements of human hands include distance features and curvature features. The distance feature represents the distance between the hand segmentation image and the center point of the palm region, which is described as follows:

$$H(\theta_q) = \max_{x_i \in A(\theta_q)} dx_i \quad (20)$$

$$f_{g,j}^h = \frac{\max_{A(\theta_{gj})} H_g(\theta)}{L_{\max}} \quad (21)$$

where $H(\theta_q)$ is the reference histogram; $A(\theta_q)$ is the angular sector θ_q of the hand corresponding to the direction; dx_i is the distance feature. Assuming that the dataset has M motions to be identified, the feature set F^h includes the distance value of each edge point $x_i \in O$ (O represents the boundary contour of the motion area) to the center C in each motion $g \in \{1, \dots, M\}$. The eigenvalue $f_{g,j}^h$ associated with the edge point x_i in action g corresponds to the maximum value of the aligned histogram. L_{\max} is the maximum distance from the edge point to the center C to ensure that all eigenvalues $f_{g,j}^h \in F^h$ are distributed within $[0, 1]$.

The curvature feature is a series of feature values obtained based on the curvature of the edge of the hand-segmented image. Consider a set of U circular dies $B_u(x_i)$ with radius r_{\max} centered on each edge sample D , taking hand edge point $x_i \in O$ and binary template $M^T(s, t)$ as input. This paper used 15 circular molds with radii ranging from 0.5 cm to 3 cm. Assuming that $V(x_i, u)$ is the curvature at x_i , the ratio of the number of samples falling into O in die $B_u(x_i)$ to the size of $B_u(x_i)$ is:

$$V(x_i, u) = \frac{\sum_{x_j \in B_u(x_i)} M(x_j)}{|B_u(x_i)|} \quad (22)$$

where $|B_u(x_i)|$ represents the base $B_u(x_i)$, $M(x_j) = M(p_j, q_j)$ represents the two-dimensional coordinate point (p_j, q_j) corresponding to x_i , and $V(x_i, u)$ needs to calculate the sample point of each $x_i \in O$, and its value range is between $[0, 1]$ (for example, $V(x_i, u) = 0.5$ corresponds to a straight edge). The $[0, 1]$ interval is quantified into N equal-sized b_1, \dots, b_N , and the corresponding value $V(x_i, u)$ of the set A of the hand edge points $x_i \in O$ falls on the mold b on u is expressed as:

$$A_{b,u} = \left\{ x_i \left| \frac{(b-1)}{M} < V(x_i, u) \leq \frac{b}{M} \right. \right\} \quad (23)$$

For each radius u and corresponding b , choose $f_{b,u}^c$ as its curvature feature, and the cardinality of the set $V(x_i, u)$ normalized by the hand edge contour length $|O|$ is:

$$f_{b,u}^c = \frac{A_{b,u}}{|O|} \quad (24)$$

The normalized curvature features are well distributed in the interval $[0, 1]$, and by sorting the gradually increasing $u = 1, 2, \dots, U$ and $b = 1, 2, \dots, N$, all the curvature features $f_{b,u}^c$ in the $U * N$ feature vector set F^c are collected as pixels with the coordinate (b, u) in the grayscale image value of. The finally obtained distance feature and curvature feature are stored in the sample feature database F_{Kinect} as the sample data F of the 3D image information.

$$F_{Kinect} = \{F^h, F^c\} \quad (25)$$

To improve the classification accuracy of complex HIMs, this paper uses the constructed multi-modal data acquisition platform for complex human hand movements to obtain sample data of time-domain signals and 3D image information of complex human movements to form a feature sample library of complex human movements.

$$F = \{F_{SEMG}, F_{Kinect}\} \quad (26)$$

5. Experimental Results

5.1. Human In-Hand Motion Recognition Results

The ten HIM recognition results based on hybrid sensors are illustrated in Figure 6, with an average recognition rate of 95.10%, reflecting a good recognition effect. The black part in the figure is the average recognition rate of each motion, and the other part is the error rate between different motions. It can be seen that the recognition rates of different motions are different. The recognition rate of motion 9 is 99%, while that of motion 4 and action 10 is only 93%. Although motions 1, 3, and 8 represent different motions, the final recognition rate is the same, which is 95%. Although the recognition effect of motion 2 and motion 6 is good, reaching the same 94%, there is still an error rate of 4% between them. The main reason may be the high similarity between the number of moving fingers and noise interference in the acquisition process. Overall, the hybrid sensor-based HIM system reveals excellent performance.

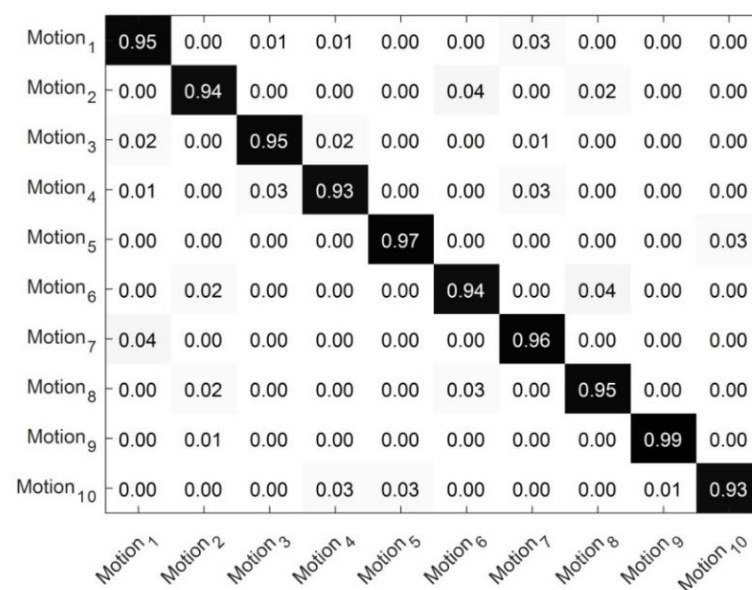


Figure 6. Confusion matrix for the ten motions using ADAG.

The main reasons for good performance are as follows: (1) Motion segmentation algorithm based on threshold ensures the segmentation of different HIMs information; (2) Multimodal data effectively guarantees the characteristics of human hand motions; and (3) A novel ADAG algorithm ensures accurate classification of the ten HIMs. However, it is also worth noting that the processing time required to classify the ten movement classes for the ten persons was 19.6 s on a Lenovo computer with an Intel(R) Core (TM) i7-4790 processor, 3.60 GHz, and 16 GB RAM with MATLAB 2018b. The processing time included the time needed to perform the motion segmentation, feature extraction, and classification of the combined data once the system had been trained.

By using different sensor-based features, the variations in the recognition rate are rather large for the same motion. In Figure 7, the recognition rate of the in-hand manipulations based on SEMG is 93.69%. Compared with Figure 6, the average recognition rate is decreased, but the recognition rate of all motions has maintained a level of over 90%. Motions 2 and 6 have the lowest recognition rate of only 92%, and their maximum error classification reaches 6%, while motions 7 and 9 have the highest recognition rate, reaching 96%. Although SEMG-based human hand gesture recognition has different degrees of misclassification, the overall recognition rate is high, which verifies the effectiveness of the feature extraction method and the machine learning algorithm.

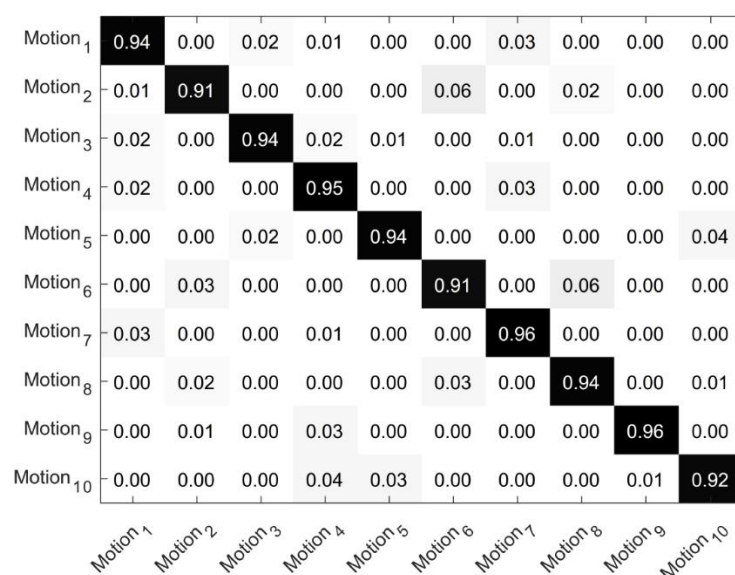


Figure 7. SEMG-based confusion matrix using ADAG.

Figure 8 indicates the confusion matrix of different motions based on Kinect. It can be clearly seen from the figure that the average recognition rate is lower, at only 89.96%. In addition, misclassification between different motions is more serious. Motion 2 has the lowest recognition rate due to its error rate of 13%. In addition, the recognition rates of motions 3, 5, 6, and 8 are all below 90%; the rest of the recognition rates are about 91%. For Kinect-based motion recognition, most researchers have adopted simple gestures or postures as the sample data for hand motion recognition in their published articles. Because the distinction of these gestures or postures is obvious, the corresponding image features are easier to distinguish, and the experimental results fully verify this view. However, in this paper, complex dynamic in-hand manipulations are used as the sample set, and the original features are less easy to acquire. Moreover, occlusion may occur due to perspective projection. Hence, these problems may be the main reasons for the Kinect-based low recognition rate.

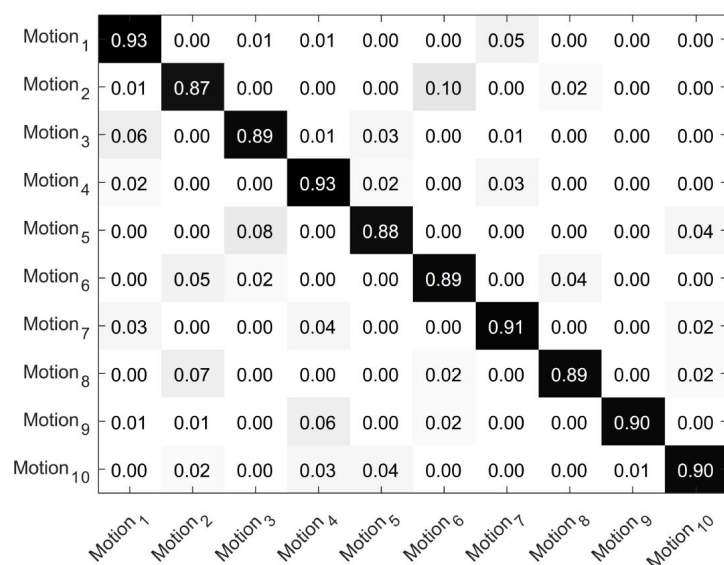


Figure 8. Kinect-based confusion matrix using ADAG.

The comparative experimental results and their variances using different sensors are shown in Figure 9. By using different sensor-based features, the variations in the recognition rate are rather large for the same motion. The Kinect-based method has the lowest average recognition rate of 89.96%, while the SEMG-based average recognition rate is 93.69%. For Kinect-based motion recognition, most researchers have adopted simple gestures or postures as the sample data for hand motion recognition in their published articles. However, in this paper, we use the complex dynamic in-hand manipulations as the sample set; the original features are less easy to acquire. Moreover, occlusion may occur due to perspective projection. Hence, these problems may be the main reasons for the Kinect-based low recognition rate. Compared with the recognition results obtained by utilizing uni-modal sensors, hybrid sensors have the highest recognition accuracy. However, the computation time is increased due to the increase in the extracted key feature information. Taken as a whole, the recognition rate distribution is relatively smooth based on different sensor types.

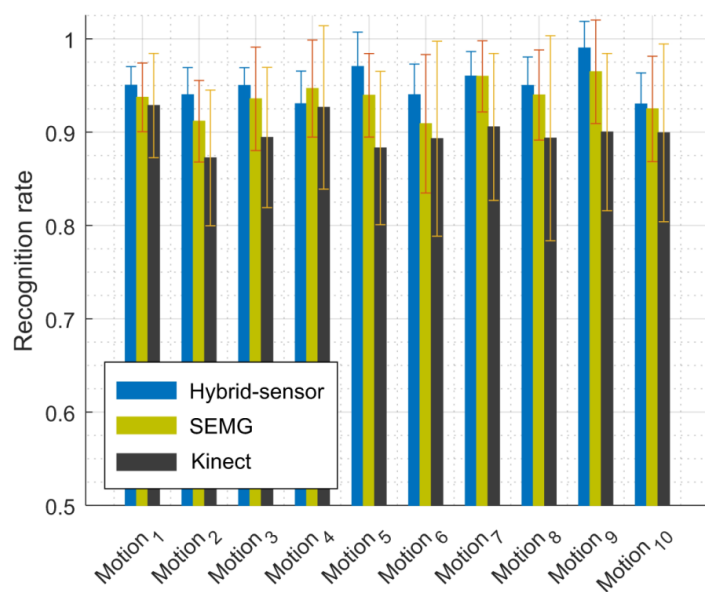


Figure 9. Comparative experimental results using ADAG.

5.2. Comparison Results of Different Subjects

Different subjects have different finger forces, SEMG signals, acceleration, etc., when completing the same motion due to their own operating habits and individual differences. Therefore, it is necessary to analyze the effects of different subjects completing the same motion on the recognition results of HIMs. The HIM recognition matrix based on different subjects is shown in Figure 10, which more intuitively reflects the recognition results of different subjects. It can be shown that different subjects have different recognition results for all HIMs. The average recognition rate of each motion reflects the overall recognition situation, while the variance represents the dispersion of different subjects for the same action result. Although there are obvious differences, the average recognition rates are as high as 92% due to the reduction of training samples and correct manipulation. Subject_1 has the highest recognition rate of over 98%, while the recognition rates of Subject_6 and Subject_8 are only 92.4%. The influence of different subjects on the final human motion recognition results cannot be ignored. In the process of data collection, more attention should be paid to the standardized training of the subjects in the manipulation.

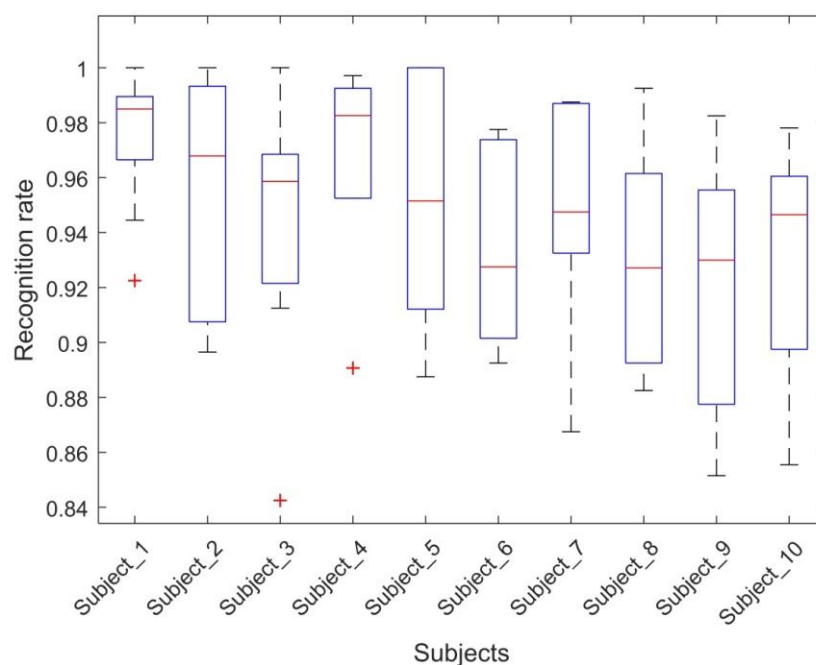


Figure 10. Recognition rates with different subjects.

5.3. Comparison Results of Different Multiclass SVM-Based Methods

In order to better verify the effect of the ADAG-SVM algorithm, the experimental results were compared with the three types of multi-class SVM methods mentioned in Table 1. Figure 11 shows the HIM recognition results obtained using four multi-class SVM algorithms. The average recognition rates based on DAG, One-Versus-One, and One-Versus-Rest are 91.52%, 94.74%, and 90.61%, respectively, which are lower than those based on the ADAG algorithm (95.10%). The similar recognition rates of One-Versus-One and ADAG are attributed to their similar basic classification principles. They both include $k(k-1)/2$ classifiers and verify all possible two-class classifiers. The One-Versus-One algorithm determines the classification of unclassified samples by voting, which has the shortcomings of false classification and refusal classification. ADAGSVM uses a decision tree structure to classify different samples, effectively avoiding the shortcomings of the One-Versus-One algorithm. It can be seen that the DAG and One-Versus-One algorithms have similar recognition rates for motion 3. Specifically, the One-Versus-One algorithm has the highest recognition rate for motion 4, while the One-Versus-Rest algorithm has the lowest recognition rates for motions 1, 3, 4, 8, 9, and 10. Overall, the ADAG algorithm

has obvious advantages over other multi-class SVM methods, and its recognition accuracy becomes more and more obvious with the number of sample classes increasing.

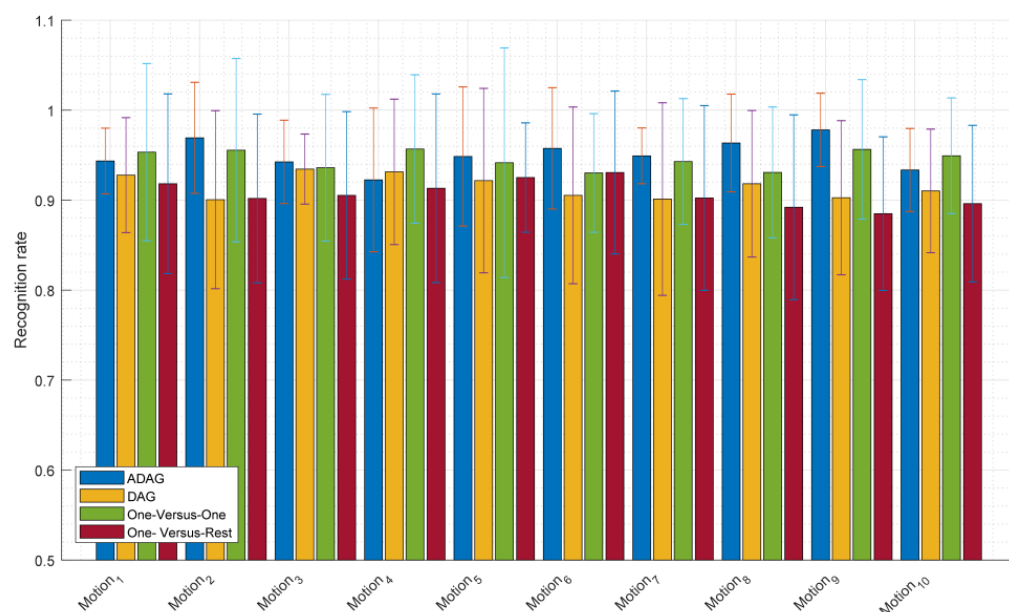


Figure 11. Recognition rates with different multi-class SVM methods.

5.4. Comparison Results of Different Machine Learning Methods

To verify that the adopted ADAG algorithm based on multi-modal sensing has more obvious advantages in human hand motion recognition than other machine learning algorithms, the K-Nearest Neighbor algorithm (KNN) and the fuzzy C-means algorithm (FCM) are selected. The KNN classifier is a machine learning algorithm that compares each feature of new data with the features used by data pairs in samples and uses distance to measure the similarity between samples. It is very suitable for sample sets that are insensitive to outliers and have no data input assumptions. As an unsupervised learning technique, the FCM algorithm has been successfully applied to feature analysis, clustering, and classifier design. Before the experiment, the parameters of the two machine learning algorithms should be set to ensure the best classification performance. The KNN algorithm takes the value $k = 5$. The FCM algorithm takes the number of clusters N , the maximum number of cycles, M , and the minimum termination cycle difference, D , as $(10, 100, 1 \times 10^{-6})$.

Figure 12 shows the single recognition results for different motions. Using different machine learning algorithms to perform multiple experiments on all motions has different recognition results. The average value reflects the overall recognition situation, and the variance represents the degree of dispersion of the recognition results for the same motion. The average recognition rate of human hand movements based on FCM is 90.16%, and the average recognition rate based on KNN is 91.57%, which is a significant gap compared with ADAG. It can be seen from the discrete degree of FCM-based recognition results that the recognition rate of different motions fluctuates greatly. Overall, the NN algorithm has obvious advantages over the other two algorithms in terms of the recognition rate of human motions.

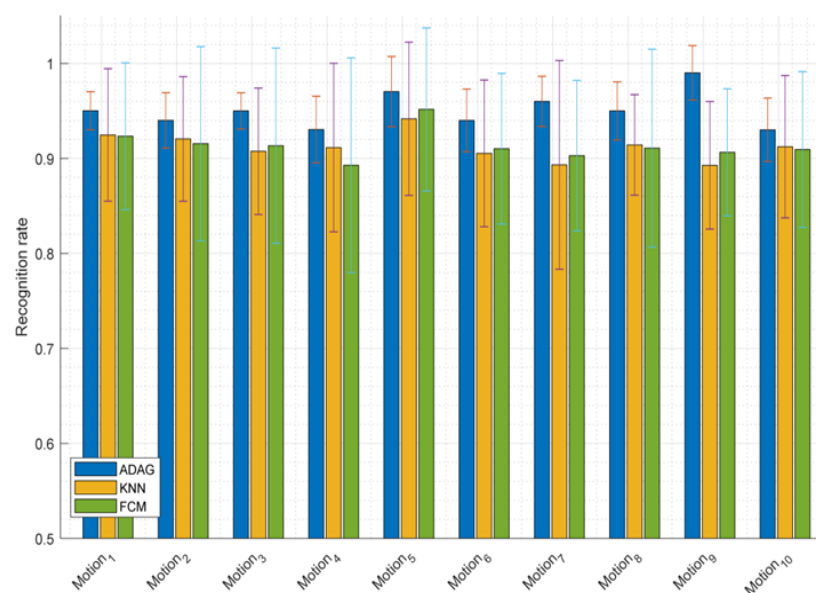


Figure 12. Recognition rates of different machine learning methods.

6. Conclusions

A hybrid SEMG- and Kinect-based HIM recognition system is proposed in this paper for robotic human-like manipulation. The system mainly consists of three parts: motion capturing, multi-modal data processing, and motion recognition. For motion capturing and data acquisition, 10 healthy adults were invited as subjects to capture HIM information, including SEMG signals and 3D scene information. For multi-modal data processing, a novel threshold-based segmentation algorithm is proposed to guarantee the validity of the raw SEMG dataset and reduce the number of repeats caused by failed motion. Six types of time-domain features from SEMG signals and distance features and hand edge features from Kinect data are extracted for HIM classification. For motion recognition, an ADAG algorithm for recognizing HIMs is proposed and optimized, and the influence of different sensors and different subjects on recognition rates is analyzed. From the experimental results, the ADAG method is compared with different sensing devices, different multi-class SVM methods, and different recognition algorithms to verify its superiority. Further research should focus on transferring the manipulation skill into prosthetic hands for human-like manipulation.

Author Contributions: Conceptualization and writing—review and editing, Y.X.; methodology, Z.J.; writing—original draft preparation, Y.X.; visualization, Y.Y., K.Y. and P.L.; supervision, K.D. and H.D.; project administration, H.D. and Y.Y.; and funding acquisition, Y.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the High-Level Talent Start-Up Fund of Pingdingshan University under Grant PXY-BSQD-2019011; in part by the Project of the Science and Technology Department of Henan Province under Grant 202102310197, Grant 212102210017, Grant 222102220116, and Grant 222102210152; in part by the Development of Robot-Enhanced therapy for children with Autism spectrum disorders of Europe FP7-ICT (DREAM) under Grant 611391.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of School of Electrical and Mechanical Engineering Academic Committee.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data used to support the findings of this study are available from the corresponding author upon request.

Acknowledgments: The authors would like to thank the reviewers for their helpful comments and constructive suggestions, which have been very useful for improving the presentation of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, T.M.; Tao, Y.; Liu, H. Current Researches and Future Development Trend of Intelligent Robot: A Review. *Int. J. Autom. Comput.* **2018**, *15*, 525–546. [[CrossRef](#)]
2. Fang, Y.; Zhou, D.; Li, K.; Liu, H. Interface prostheses with classifier-feedback-based user training. *IEEE Trans. Biomed. Eng.* **2017**, *64*, 2575–2583. [[PubMed](#)]
3. Li, G.; Ma, F.; Guo, J.; Zhao, H. Bringing robotics to formal education: The Thymio open-source hardware robot. *IEEE Robot. Autom. Mag.* **2017**, *24*, 77–85.
4. Wang, L.; Hu, W.; Tan, T. Recent developments in human motion analysis. *Pattern Recognit.* **2017**, *36*, 585–601. [[CrossRef](#)]
5. Xue, Y.; Ju, Z.; Xiang, K.; Chen, J.; Liu, H. Multiple sensors based hand motion recognition using adaptive directed acyclic graph. *Appl. Sci.* **2017**, *7*, 358. [[CrossRef](#)]
6. Vishwakarma, D.K. Hand gesture recognition using shape and texture evidences in complex background. In Proceedings of the 2017 International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, India, 23–24 November 2017; pp. 278–283.
7. Xue, Y.; Ju, Z.; Xiang, K.; Chen, J.; Liu, H. Multimodal Human Hand Motion Sensing and Analysis—A Review. *IEEE Trans. Cogn. Dev. Syst.* **2019**, *11*, 162–175.
8. Elliott, J.M.; Connolly, K. A classification of manipulative hand movements. *Dev. Med. Child Neurol.* **1984**, *26*, 283–296. [[CrossRef](#)] [[PubMed](#)]
9. Exner, C.E. The zone of proximal development in in-hand manipulation skills of nondysfunctional 3-and 4-year-old children. *Am. J. Occup. Ther.* **1990**, *44*, 884–891. [[CrossRef](#)] [[PubMed](#)]
10. Pont, K.; Wallen, M.; Bundy, A. Conceptualising a modified system for classification of in-hand manipulation. *Aust. Occup. Ther. J.* **2009**, *56*, 2–15. [[CrossRef](#)] [[PubMed](#)]
11. Xue, Y.; Yu, Y.; Yin, K.; Li, P.; Xie, S.; Ju, Z. Human In-Hand Motion Recognition Based on Multi-Modal Perception Information Fusion. *IEEE Sens. J.* **2022**, *22*, 6793–6805. [[CrossRef](#)]
12. IBullock, M.; Ma, R.R.; Dollar, A.M. A hand-centric classification of human and robot dexterous manipulation. *IEEE Trans. Haptics* **2013**, *6*, 129–144. [[CrossRef](#)] [[PubMed](#)]
13. Andrychowicz, O.M.; Baker, B.; Chociej, M.; Józefowicz, R.; McGrew, B.; Pachocki, J.; Petron, A.; Plappert, M.; Powell, G.; Ray, A.; et al. Learning dexterous in-hand manipulation. *Int. J. Robot. Res.* **2020**, *39*, 3–20. [[CrossRef](#)]
14. Pagoli, A.; Chapelle, F.; Corrales, J.A.; Mezouar, Y.; Lapusta, Y. A soft robotic gripper with an active palm and reconfigurable fingers for fully dexterous in-hand manipulation. *IEEE Robot. Autom. Lett.* **2021**, *6*, 7706–7713. [[CrossRef](#)]
15. Li, K.; Zhang, J.; Wang, L.; Zhang, M.; Li, J.; Bao, S. A review of the key technologies for sEMG-based human-robot interaction systems. *Biomed. Signal Process. Control* **2020**, *62*, 102074. [[CrossRef](#)]
16. Xu, H.; Xiong, A. Advances and Disturbances in sEMG-Based Intentions and Movements Recognition: A Review. *IEEE Sens. J.* **2021**, *21*, 13019–13028. [[CrossRef](#)]
17. Hu, Y.; Wong, Y.; Wei, W.; Du, Y.; Kankanhalli, M.; Geng, W. A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition. *PLoS ONE* **2018**, *13*, e0206049. [[CrossRef](#)] [[PubMed](#)]
18. Ma, R.; Zhang, L.; Li, G.; Jiang, D.; Xu, S.; Chen, D. Grasping force prediction based on sEMG signals. *Alex. Eng. J.* **2020**, *59*, 1135–1147. [[CrossRef](#)]
19. Wang, L.; Huynh, D.Q.; Koniusz, P. A Comparative Review of Recent Kinect-Based Action Recognition Algorithms. *IEEE Trans. Image Process.* **2020**, *29*, 15–28. [[CrossRef](#)] [[PubMed](#)]
20. Ju, Z.; Ouyang, G.; Liu, H. EMG-EMG correlation analysis for human hand movements. In Proceedings of the 2013 IEEE Workshop on Robotic Intelligence in Informationally Structured Space (RiIS), Singapore, 16–19 April 2013; pp. 38–42.
21. Sun, J.; Wang, Y.; Li, J.; Wan, W.; Cheng, D.; Zhang, H. View-invariant gait recognition based on kinect skeleton feature. *Multimed. Tools Appl.* **2018**, *77*, 24909–24935. [[CrossRef](#)]
22. Chauhan, V.K.; Dahiya, K.; Sharma, A. Problem formulations and solvers in linear SVM: A review. *Artif. Intell. Rev.* **2019**, *52*, 803–855. [[CrossRef](#)]
23. Hsu, C.-W.; Lin, C.-J. A comparison of methods for multiclass support vector machines. *IEEE Trans. Neural Netw.* **2002**, *13*, 415–425. [[PubMed](#)]
24. Tkach, D.; Huang, H.; Kuiken, T.A. Study of stability of time domain features for electromyographic pattern recognition. *J. Neuroeng. Rehabil.* **2010**, *7*, 21. [[CrossRef](#)] [[PubMed](#)]