Multimedia Tools and Applications
https://doi.org/10.1007/s11042-023-15252-w

Ultrasonic positioning and IMU data fusion for pen-based 3D hand gesture recognition

Siyu Liu[1] · Jian Chen[1] · Cheng Wang[1] · Lin Lin[1]

Received: 10 August 2021 / Revised: 9 December 2022 / Accepted: 6 April 2023

Abstract
In this paper, a pen-based 3D hand gesture dataset and recognition method using ultrasonic positioning and inertial data is proposed. First, considering that 3D hand gestures have six degrees of freedom, a 3D hand gesture dataset based on trajectory shape attributes, motion direction attributes and pen attitude attributes is proposed. Then, each attribute of the gesture is processed according to its priority, and the corresponding data channel and recognition method are selected to determine the 3D hand gesture label. Finally, experimental verification is conducted using a 3D multi-channel pen-like interactive device. For a 10-gesture set, the gesture recognition rates achieved ranged from 86.5–99.5%, depending on whether a single or multiple templates and thresholds are used. The results show that the 3D hand gesture recognition method proposed in this paper can recognize pen-based gestures effectively and solve the problem of traditional gesture recognition methods not being able to recognize 3D hand gestures containing multiple attributes.

Keywords Human–computer interaction · 3D hand gesture recognition · 3D positioning · 3D pen-like interaction · Inertial sensor · Multichannel interaction

1 Introduction

With the development of computer technology, the usability of commercial human–computer interactions (HCI) has improved considerably over the past decades. An enormous amount of research effort goes into this area to answer an increasingly complexity of interactive scenes [26]. Hand gesture gives a natural, non-verbal form of communication that can replace or reinforce other communication modalities such as speech or writing. Comparing to the classical surface contacting hand gesture, i.e., keyboard and mouse, three-dimensional hand gesture (3DHG) provide an intuitive expression of intentions of people in a more natural and flexible manner. Nowadays, 3DHG interaction has been the mainstream

✉ Jian Chen
chenjian@jlu.edu.cn

[1] College of Communication Engineering, Jilin University, Changchun 130022, China

technique for HCI and emerged as a major means of interaction in games, augmented reality (AR), virtual reality (VR) platforms and other related applications [19, 26, 30, 31].

Hand gesture recognition is a challenging problem which have attracted a lot of attention of investigators. Three basic types of sensors have so far been commonly used for sensing hand gesture: multitouch screen sensors, vision-based sensors and mount-based sensors [11]. In the first case, touch gestures are identified by keeping track of fingers which restricting the diversity of operating command. It is especially suitable for mobile devices due to the intuitively contacting between user and devices [27]. Vision-based hand gestures recognition usually employs traditional video/image or depth camera combined machine learning to capture hand gesture. Many previous works focus on the image information such as the shape, color of hand gestures, and yet such information is sensitive to illumination and complicated background [1, 5, 7, 14, 17, 20, 21, 32]. Moreover, the emergence and rapid development of depth cameras provide a new method solve this task. A depth image contains information of the distances of all the pixels from the surface of the object to the sensors so that 3D information collection is hardly affected by cluttered background or light changing. For instance, Zhu et al. suggested a hand gesture recognition method based on depth images captured via the Kinetic sensors [33]. A new 3D shape descriptor is proposed for extracting the pattern of salient 3D shape features. However, the artificial parameters should be preset to make tradeoff between the accuracy and cost. Furthermore, the approaches based on depth images frequently encounter a large amount of contour noise caused by fast hand motion. Rotation and global position of hand may result in the reduced accuracy of classification algorithm [7].

Meanwhile, a large number of studies concentrate on hand gesture recognition using mount-based sensors, i.e., surface electromyography sensors and inertial sensors [24, 25]. Pezzuoli et al. investigated the recognition and classification of 27 dynamic hand gestures by different neural network and machine learning variants. A fairly high classification accuracy has been achieved. However, the sensors for data collection are integrated in data-glove may lead to the loss of flexibility and comfort level [23]. Embedding an inertial sensor, such as accelerometer or gyroscope, in a pen-type input device can accurately capture the acceleration and angular velocity of hand movement, and be free from the restrict of interactive space and hand gestures. Then, the hand gestures can be classified with the data retrieved from these sensors. Therefore, the pen-based interaction accumulates inertial sensors have been widely exploited in hand gesture recognition [2, 12, 15, 16, 29]. Akan [2] implemented a gesture recognition system with a neural network and multi-sensors. A mobile phone is used to obtain the 3-axis measurements of the accelerometer, magnetometer, gyroscope and orientation. The performance of their approach in hand gestures recognition is user-dependent, ranging from 60% to 91.66%. Hsu [15] used an inertial pen for English lowercase letter recognition and the Dynamic Time Warping (DTW) algorithm was adopted for handwriting recognition. There will be fairly large error when estimating the position of the pen due to the limitations of single inertial sensor, and this is the common problem for the investigations which using a single sensor channel to acquire the features of the 3DHGs [3, 4, 13]. Some studies make constraints on the size or position of 3DHGs and the trajectory of hand gestures should be consistent during interaction as far as possible [8, 18]. So that the comfort and naturality of the interaction is restricted to a great degree. Actually, a real 3DHG possesses higher degree of freedom in the practical operations. The 3DHGs with the same trajectory shape may have different motion direction or rotational situations in 3D space, which can be endowed with different functions to express different intensions [22]. The efficiency and flexibility of HCI will be augmented while all these attributes of hand gestures are taken into consideration. However, it is still a

great challenge to fulfill high precision for hand gestures recognition with higher degree of freedom. Most previous research only focuses on the trajectory shape of hand gestures and capture the trajectory feature for the following recognition.

As mentioned above, the accuracy of hand gestures recognition of many studies is high enough, but the work considers both multi-channel sensor data and multiple attributes of hand gestures is far from complete. For that the main contribution of this paper is to fill this gap and the main innovations are: (*i*) a new 3DHGs dataset based on trajectory shape, direction of motion and pen attitude is proposed for captures the comprehensive information of 3DHGs; (*ii*) multi-channel data i.e., acceleration and attitude, are adopted to describe 3DHG attributes which is helpful in improving recognition rate.

The rest of this paper is organized as follows. In Sections II, a blend of multiple attributes, i.e., trajectory shape, movement direction and pen attitude, is used for 3DHGs definition a 3DHG dataset is defined subsequently. The 3DHG recognition algorithm consists of trajectory shape recognition, motion recognition of the pen tip and pen body, and pen attitude recognition is presented in Section III. In Section IV, the detailed experimental process is described, followed by the results and conclusion are given in the last two sections.

2 Multi-attributes of 3DHGs and dataset

To implement 3D pen-like interaction with multiple features and contain more additional information for complex interaction scenarios, 3DHGs must be composed of the trajectory shape, the direction of motion, and the pen attitude. These attributes will form the building blocks for 3DHGs recognition. On this basis, multi-channel data such as spatial coordinates, acceleration and orientation are fused to describe different attributes of 3DHGs. The attributes and 3DHGs dataset are described in more detail in the subsequent chapters.

2.1 Trajectory shape

For a pen-like interaction, the spatial trajectory of pen in the interactive process is considered as the trajectory shape of a 3DHGs idiomatically. However, users may encounter the memory difficulties if the 3DHGs is impenetrable and complicated, higher learning cost is required. Learnable and easy-to-remember hand gestures are more suitable for experimentation. Studies have shown that individuals' most suitable working memory capacity for hand gestures is only 3 to 4 [6]. Therefore, as shown in Fig. 1, three trajectory primitives are used in this work, namely the line segment, the circle and the V-shape, which respectively represent line segments, closed geometric figures and polylines. These three primitives refer only to the trajectory shape of the pen in the hand gesture action.

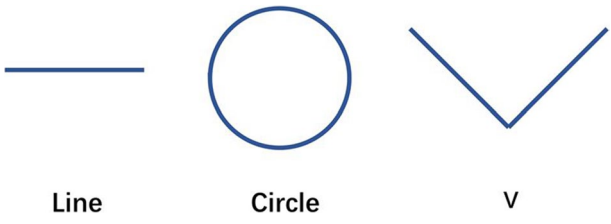


Fig. 1 Schematic diagram of primitives

Moreover, previous studies have made request for keeping the size of the trajectory as close as possible to the reference values or the position and rotation of 3DHG should be consistent during the collection of hand gesture data. However, it is quite difficult to meet the synchronization requirements in practical applications [8, 18]. Actually, the uniqueness and exactitude of the interactive intension for a specific 3DHG means that it should be immune from the size, position and rotation of trajectory. Therefore, the above-mentioned characteristics of trajectories are fully considered in present work. The size, position and rotation of trajectories can be transformed arbitrarily for a certain 3DHG which have the same interactive intensions. For circle and V primitives, there is also no restriction on the spatial orientation.

2.2 Motion direction

The motion direction of the tip and / or the body for the electronic pen relative to an external reference system during the hand gesture action is defined as the direction of motion of the 3DHGs. Specifically, the motion direction of the pen tip is relative to the reference coordinate system in 3D space, where X, Y, and Z are the axes of the reference coordinate system, shown in Fig. 2. Meanwhile, as can be seen in Fig. 3, the coordinate axes X_1, Y_1, Z_1 of the three-axis acceleration sensor in the pen are used as the reference coordinate system for the motion of the pen's body. Therefore, motion direction of the pen body is affected by the orientation of the pen. With this treatment, the corresponding relationship between the movement of pen tip and pen body during interaction can be obtained and it can give a more precise description for 3DHGs. It enables us to hold a pen in a more flexible way and can also expand the diversity of interactive intensions.

In the present work, for a gesture trajectory consisting of a line segment, 6 different motion directions for pen tip and pen body along the axes of corresponding reference frame have been defined, respectively. The different combination of motion directions of pen tip and pen body can express different intentions in interactive process. For simplify, the motion direction of pen tip and pen body are not considered in circular and V-shaped trajectory gestures.

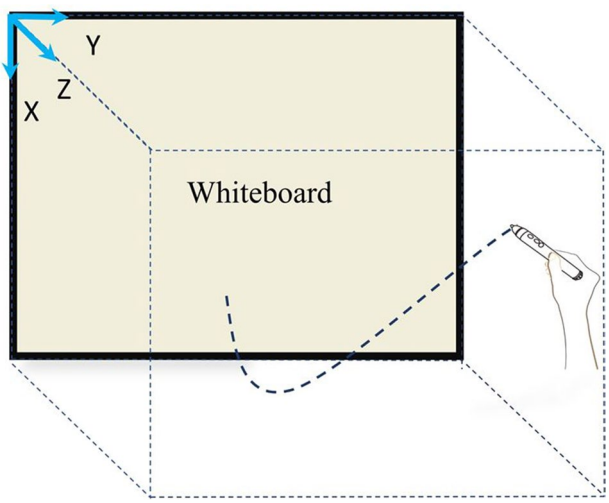


Fig. 2 Reference frame for the direction of motion of the pen tip

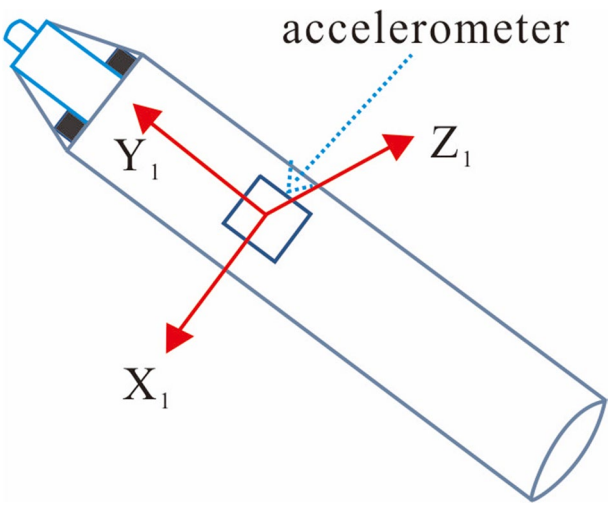


Fig. 3 Reference frame for the motion of the pen body

2.3 Pen Attitude

The pen attitude during different hand gestures is a feature of the basic interactive intention expressed by the trajectory's morphological attributes. Due to the characteristics of the gesture trajectory, during line-segment gestures the general pen attitude remains relatively unchanged as the action is performed. If the relative attitude of the pen changes, the trajectory shape obtained is a polyline or a curve. For circular and V-shaped gestures, two possible pen attitudes are defined to complete these gestures. One is to maintain the pen body fixed and complete the hand gesture through translation; the other is to complete the gesture by rotating the wrist and the pen body, as shown in Fig. 4.

2.4 3D hand gesture dataset

Based on aforementioned attributes, ten typical hand gestures for 3D pen-like interactions are defined and summarized in Table 1. The gestures corresponding to labels 1 and 3 are shown schematically in Fig. 5, where the green arrow and the red arrow indicate the motion direction of the pen tip and pen body, respectively. The combination of motion direction of pen tip and pen body are considered in line segment and pen attitude is considered in circular and V-shaped gestures. Moreover, the 3D position coordinate sequence is used to

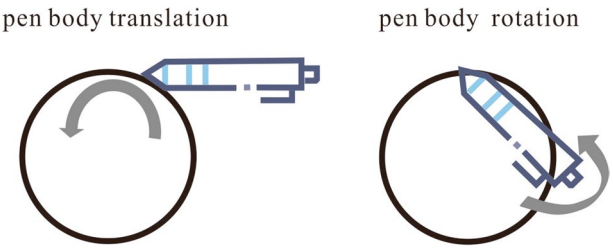


Fig. 4 Schematic diagram of pen body translation and rotation in 3D space

Table 1 Overview of 3DHG sets

Label	1	2	3	4	5	6	7	8	9	10
Trajectory shape	Line	Line	Line	Line	Line	Line	Circular	Circular	V-shaped	V-shaped
Motion direction	$+X$ $+X_1$	$-X$ $-X_1$	$+Y$ $+Z_1$	$-Y$ $-Z_1$	$+Z$ $-Y_1$	$-Z_1$ $+Y_1$	Random	Random	Random	Random
Pen attitude	Random	Random	Random	Random	Random	Random	Translation	Rotation	Translation	Rotation

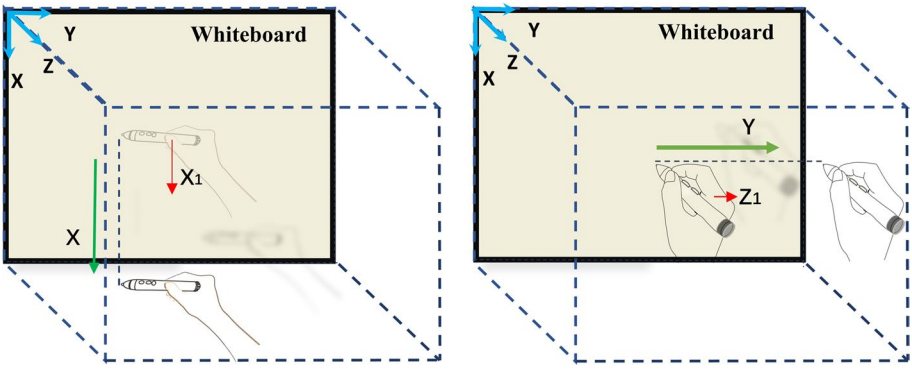


Fig. 5 Schematic diagram of motion direction of pen tip and pen body for hand gestures (**a**) label 1, (**b**) label 3

illuminate the pen movement trajectory during the interactive process. The quaternions are utilized to calculate and characterize the information contained in the pen's attitude.

3 3D hand gesture recognition

The 3DHG attributes are assessed according to their priority and the appropriate data channel was determined based on the assessment. The trajectory shape is the basic attribute of all 3DHGs, so it is analyzed first. The recognition will be completed according to the motion direction or pen attitude subsequently. This detail process of 3DHGs recognition is shown schematically in Fig. 6.

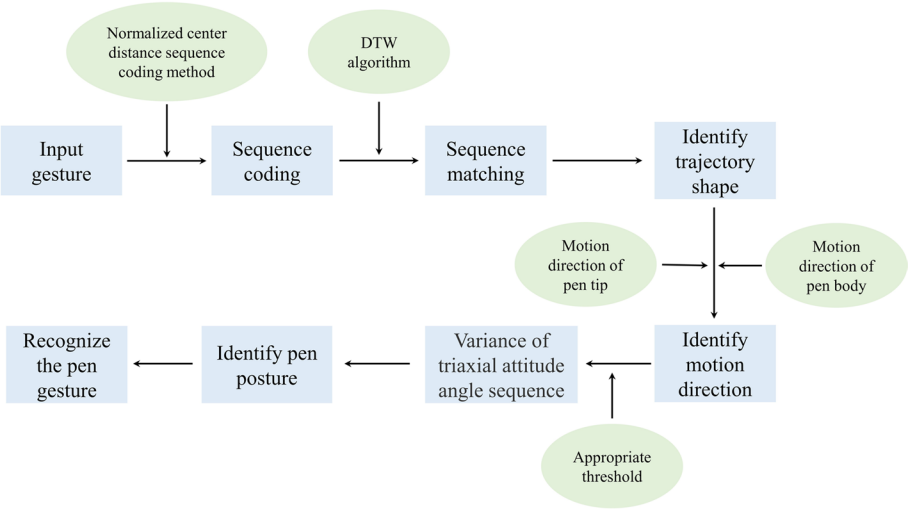


Fig. 6 Schematic diagram of 3DHGs recognition method

3.1 Trajectory shape recognition

A trajectory encoding based on the normalized center distance is used to characterize the morphological characteristics of the 3DHGs trajectory. A 3D coordinate sequence is required and then the corresponding normalized center distance coding sequence will be obtained subsequently. This method can simultaneously satisfy the requirement of that change of the size, position and rotation of trajectory have no influence on the 3DHGs recognition. The specific process of the trajectory recognition is as follows:

The invalid data are removed at pre-processing process and then the non-resampled 3D coordinate sequence of the pen tip is used to calculate the 3D coordinates of the center of trajectory $c_{\mathrm{c}} = (x_{\mathrm{c}} + y_{\mathrm{c}} + z_{\mathrm{c}})$,

$$c_{\mathrm{c}} = \frac{1}{M} \sum_{i=1}^{M} c_{\mathrm{i}} \tag{1}$$

where c_i is the 3D position coordinates of the pen tip for the i^{th} sample and M is the sequence length. The coding sequence $d[m] = d_1, d_2, \ldots d_M$ is then obtained by calculating the Euclidean distance from each set of position coordinates c_i to the trajectory center c_c. The coding sequence is invariant to translational motion and rotation obviously. Then, the linear distances between two adjacent points are summed to obtain the full length of the trajectory and normalize the coding sequence of the trajectory center distance $d[m]$. In this manner, the normalized trajectory center distance coding sequence can be calculated by $s[m] = 1/L \times d[m]$. Therefore, the size of the trajectories will not affect the recognition of 3DHGs.

In practical interaction scenarios, the interaction duration of different gestures is usually inconsistent, and it may result in different code sequence lengths for each sample. In the present work, DTW is used for identifying the coding sequences of samples by measuring the similarity between different time sequences [15]. Template sequences corresponding to different trajectory shapes are compared to the sample sequence, and the template label with the smallest distance is the recognition outcome for this particular sample. The specific steps for creating the template sequence are as follows:

(i) In a preliminary experiment, participant i collects p sets of different trajectories. The lengths of the coding sequences are $l_1, l_2, \ldots, l_p$, respectively. Then, the total length of the template sequence for the participant is:

$$L_i = \frac{1}{p} \sum_{k=1}^{p} l_k \tag{2}$$

(ii) In multiple templates and thresholds (MTT) experiment with q participants, every participant generates a corresponding template length L_i ($i = 1, 2, \ldots, q$), while an average template length L_{ave} is used for gesture recognition in single template and threshold (STT) experiments:

$$L_{ave} = \frac{1}{q} \sum_{k=1}^{q} L_k \tag{3}$$

(iii) The normalized coding template sequences of the trajectory center distances for the line segment, circular and shape trajectories are obtained based on the tem-

plate length and trajectory shapes, which are respectively denoted as T_1, T_2, T_3. Then, the encoded sample and template sequences are compared using the DTW. The template and sample sequences are denoted as T_1 and S, with lengths $|T_1|$ and $|S|$, respectively. An optimal warp path for two sequences can be expressed as $W = w_1, w_2, ..., w_k, \max(|T_1|, |S|) \leq K \leq (|T_1| + |S|)$, where K is the length of the warp path and the k^{th} element of the warp path is $w_k = (i, j)$. The minimal distance for the whole warped path is calculated dynamically using:

$$D(i,j) = D_{\text{ist}}(i,j) + min[D(i-1,j), D(i,j-1), D(i-1,j-1)] \tag{4}$$

$$D_{ist}(i,j) = \sqrt{(T_1[i] - S[j])^2} \tag{5}$$

where $D_{\text{ist}}(i, j)$ is the Euclidean distance between $T_1[i]$ and $S[j]$.

Then, the shortest warped path distance represents the similarity between the sample sequence and the template sequence:

$$D(|T_1|, |S|) = \text{argmin}(\sum_{t=1}^{k} w_t(i(t), j(t))) \tag{6}$$

For an encoded sample sequence to be identified, DTW matching is performed with the template sequence of the three different trajectories T_1, T_2, T_3, and three warped distances D_1, D_2, D_3 are obtained. The trajectory of the sample is matched to the template trajectory with the smallest warped distance.

3.2 Motion direction recognition

The 3D coordinate sequences of the hand trajectory are used to extract the movement of pen tip. The angle between the pen tip motion trajectory and each coordinate axis is estimated to identify the pen tip motion direction. The corresponding steps are:

(i) The 3D coordinate sequence of each sample is divided equally into r intervals and the coordinates of the end and start points are subtracted to obtain the direction vector $v_1, v_2, \ldots, v_k$ for every interval.

(ii) Then, the angles between the direction vector and the unit vectors on the X-, Y- and Z-axis directions are calculated within a range of $[0, \pi/2]$). α, β, γ are defined as the average values of the angle relative to each axis. Then, the motion direction of the pen tip is determined as being along the axis corresponding to the minimum of (α, β, γ). The sign of the difference between an interval's start and end point coordinates denotes the motion direction of trajectory relative to the corresponding positive or negative axis direction.

The three-axis acceleration sequence is used to estimate the three-axis movement speed of the pen body during the gesture and further used to identify the motion direction of the pen body. The velocity component sequence of the electronic pen along the direction axis can be calculated using the acceleration sequence along the same axis and the time interval between two data points. The mean value of this sequence is the average velocity:

$$v_{\text{ave}} = \frac{1}{N}\sum_{i=1}^{N} v_i \tag{7}$$

where N is the sequence length of hand gesture.

Finally, the motion direction of pen body can be determined by comparing the modes of the average speed. The pen moves along the i^{th} axis while i corresponds the axis with maximum ($|vx_{\text{ave}}|$, $|vy_{\text{ave}}|$, $|vz_{\text{ave}}|$), where $|vx_{\text{ave}}|$, $|vy_{\text{ave}}|$, $|vz_{\text{ave}}|$ are the absolute values of the average velocity along the X_1, Y_1, Z_1 axes. $v_{ave} \geq 0$ refers to the pen moving along the axis in the positive direction while $v_{\text{ave}} < 0$ refers to the negative direction.

3.3 Pen attitude recognition

The change of the pen attitude may take place during the translation or rotation of the pen in a 3DHGs set. The change degree of the pen attitude can be determined by calculating the variance of the attitude angle sequence during the gesture.

First, the quaternion is collected using the inertial sensor and further converted into an attitude angle. The quaternion corresponding to the hand gesture is expressed as $q = q_0 + q_1 i + q_2 j + q_3 k$. The roll angle ϕ_{i}, pitch angle θ_{i} and yaw angle ψ_{i} (Fig. 7) of the pen read during the i^{th} sampling period are extracted via the quaternions q_{i0}, q_{i1}, q_{i2}, q_{i3}, respectively, using the following relationship:

$$\begin{pmatrix} \phi_i \\ \theta_i \\ \psi_i \end{pmatrix} = \begin{pmatrix} arctan\dfrac{2(q_{i0}q_{i1} + q_{i2}q_{i3})}{1 - 2(q_{i1}^2 + q_{i2}^2)} \\ arcsin 2(q_{i0}q_{i2} + q_{i1}q_{i3}) \\ arctan\dfrac{2(q_{i0}q_{i3} + q_{i1}q_{i2})}{1 - 2(q_{i2}^2 + q_{i3}^2)} \end{pmatrix} \tag{8}$$

The roll, pitch and yaw angle sequences for the pen body attitude in the gesture process can be represented as $\phi[n] = \phi_1, \phi_2, ..., \phi_N, \theta[n] = \theta_1, \theta_2, ..., \theta_N$ and $\psi[n] = \psi_1, \psi_2, \ldots, \psi_N$, respectively. Then the variance of the attitude angle sequence in each case is calculated to obtain V_ϕ, V_θ, V_ψ. The algebraic average value of the three angles $V_{\text{ave}} = \left(V_\phi + V_\theta + V_\psi\right)/3$ is used to reflect the change degree of the pen attitude during the hand gesture process.

In pre-experiment, the translation and the rotation of the pen are be processed separately. The algebraic average angle value of translation and the rotation V_{ave}^t, V_{ave}^r are calculated first. Then the sequence of translational and rotational samples is recorded as $V_{ave}^t[i]$, $V_{ave}^t[i]$, respectively. The average change of the pen attitude for translational and rotational samples are denoted as V_{t} and V_{r}, which can be yield by $V_{t(r)} = \frac{1}{m}\sum_{i=1}^{m} V_{ave}^{t(i)}[i]$, where

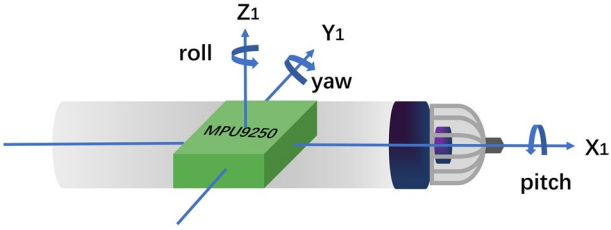


Fig. 7 Attitude angle of the electronic pen

m is the number of translational or rotational samples. An appropriate threshold, expressed as $V_{th} = \left(V_t + V_r\right)/2$, is used to distinguish the change of the pen's attitude. If $V_{ave} \geq V_{\text{th}}$, the pen body is performing translational motion during the gesture, while if $V_{ave} < V_{\text{th}}$, the pen body is being rotated.

Pen body attitude screening thresholds are also divided into single threshold and multiple thresholds. Single thresholds are the arithmetic mean value of all experimenters' thresholds, while the latter one will calculate the threshold for each participant according to the above-mentioned method.

4 Experiments

A 3D ultrasonic pen-like interaction device including a host computer and an electronic pen was used in the experiment. As shown in Fig. 8, the host included 6 positioning receivers: R1, R2, …, R6. The serial peripheral interface (SPI) communication protocol was used

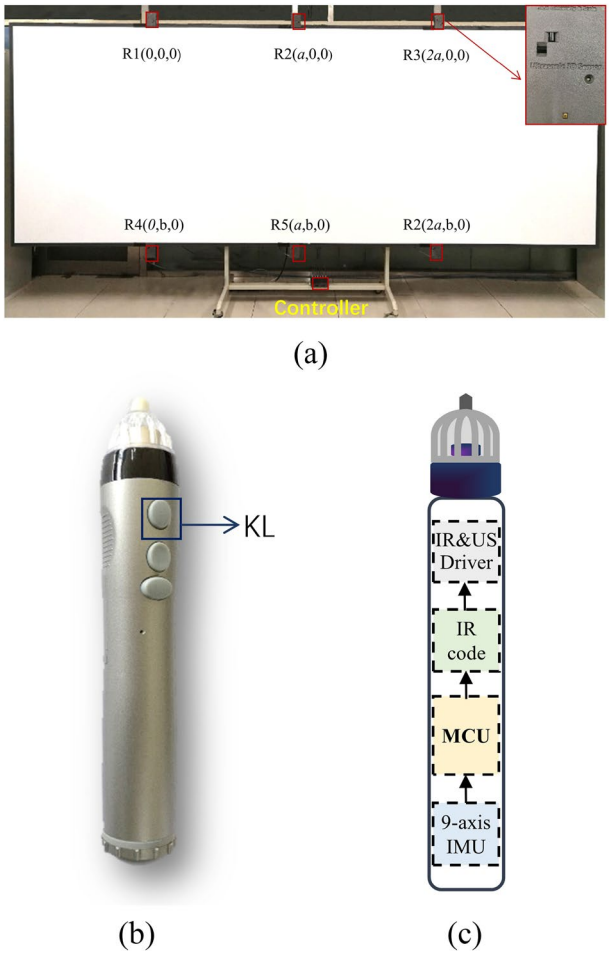


Fig. 8 3DHG interaction devices. (**a**) Host computer; (**b**) External keys of electronic pen, (**c**) Internal structure of electronic pen

for the communication between the controller and receiver. On this basis, the 3D coordinates of the electronic pen can be obtained in real time [9, 10].

6 ultrasonic receivers are placed on the upper and lower edges of a whiteboard on the XY-plane (Fig. 8(a)). The detection range of the device was 1.5 m × 4.0 m × 1.5 m, and the position accuracy was 2 mm. The electronic pen is composed of an MCU, an infrared (IR) driving circuit (driver) and an IR transmitter, an ultrasound (US) driver, a Polyvinylidene Fluoride (PVDF) US transmitter, and a 9-axis MPU9250 inertial measurement unit (IMU). The PVDF piezoelectric film was a US transmitter, which emitted the US signals required for positioning. The IR emission tube emitted the reference signal for US timing. The wireless communication chip was responsible for the data transmission between the electronic pen and the host computer. There are four buttons on the electronic pen. In the present work, the start and end of gestures are determined by triggering of the KL button (Fig. 8(b)).

Ten dominantly right-handed participants aged 22–25 years conduct a preliminary experiment to determine the length of the template sequence used in the DTW algorithm and the pen body attitude thresholds. In MTT experiments, each participant had his own template and threshold, while in STT experiments, all participants shared a set of template and threshold obtained by averaging pre-experiment values. Each participant performed 10 types of hand gestures for 20 times in the experimental, respectively. The detailed data collection process was conducted as follows.

The pen is powered on and the initial attitude was determined at the initial stage of the experiment. The participant holds the electronic pen in a normal holding position and then the participant pressed the KL button on the electronic pen to start a specific hand gesture interaction. The Unity 3D platform is used to collect gesture data and store the original data of the US and IMU modules. The interactive interface in Unity 3D is shown in Fig. 9. The real-time trajectory will show in the interface and can be deleted by clicking the "Clear"

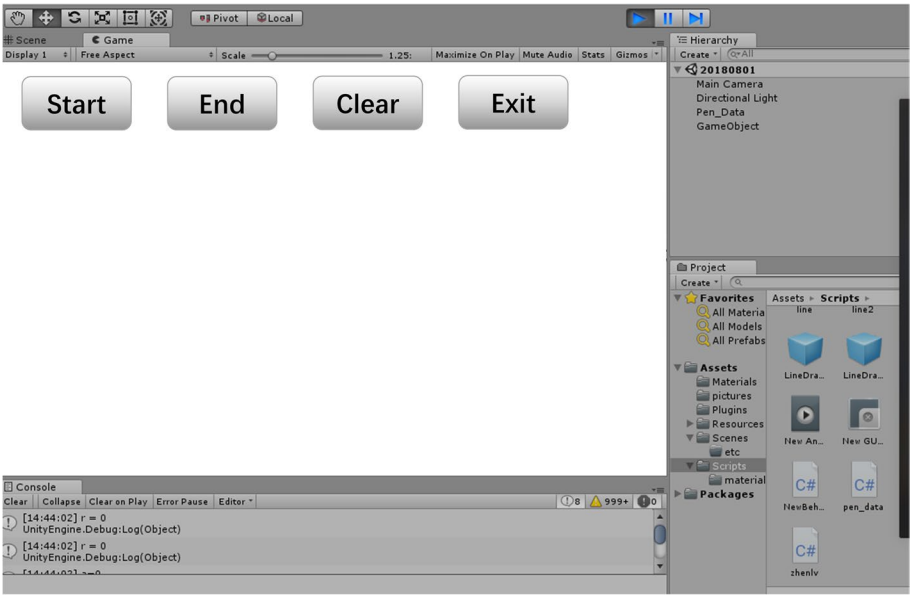


Fig. 9 Hand gesture data acquisition interactive interface in Unity 3D

Table 2 Number of correct recognitions of each label for different participants in MTT experiments

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10
Num 1	20	20	19	20	20	17	20	20	19	20
Num 2	20	20	17	20	20	20	20	20	19	18
Num 3	19	20	20	20	20	17	20	20	20	20
Num 4	20	20	20	19	20	18	20	20	20	18
Num 5	20	20	20	20	17	20	20	19	16	20
Num 6	20	20	16	17	20	16	20	20	19	20
Num 7	20	20	20	16	17	20	20	20	20	19
Num 8	19	20	20	20	20	20	18	17	20	20
Num 9	20	19	20	20	20	20	20	20	18	20
Num 10	19	20	20	20	20	20	20	20	17	18

button. Clicking the "Exit" button to exit the data collection process. The output function of the Unity 3D environment is used to save the data into a specific location for subsequent processing. The gesture completed once the participant releases the KL button.

5 Results and discussion

The quantity of correctly identified samples of different 3DHGs for each participant is summarized in Table 2–3. The recognition rate of each 3DHG label refers to the percentage of correctly identified samples corresponding to that label for this specific experiment. A comparison of the recognition rates for each hand gesture in STT and MTT experiments is shown in Fig. 10. It can be found that all the recognition rates of different 3DHGs are larger than 94%. Correspondingly, the recognition rates for different 3DHGs in STT experiment are all reduced to a different extend. Labels 8 and 9 had the greatest decrease in recognition rate (7.5% each), while labels 3 and 6 had the least decline in recognition rates (4%).

The relatively lower recognition rate of STT experiment can be ascribed to a common template and threshold is shared in the recognition process for all participants. The hand gestures will differ slightly between different participants so that a shared template and threshold may affect the recognition of a specific hand gesture. Moreover, the recognition rates of labels 9 and 10 are fairly low in both MTT and STT experiments. The most essential

Table 3 Number of correct recognitions of each label for different participants in STT experiments

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10
Num 1	19	20	18	16	20	17	20	17	16	19
Num 2	18	19	16	17	20	20	20	17	18	17
Num 3	20	18	20	17	19	20	17	16	20	18
Num 4	19	18	20	20	16	17	17	20	19	18
Num 5	20	19	20	20	20	18	20	17	18	20
Num 6	19	20	18	16	17	18	20	18	17	19
Num 7	16	16	17	18	17	16	16	20	14	19
Num 8	20	18	19	20	18	16	17	19	16	16
Num 9	19	20	20	18	17	20	20	18	18	18
Num 10	18	20	16	17	18	19	20	20	17	18

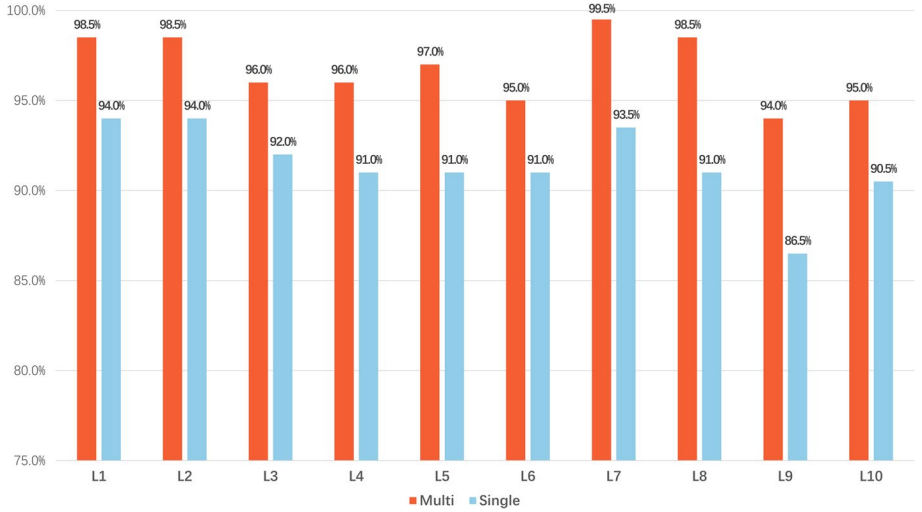


Fig. 10 Comparison of label recognition rates

reason for this phenomenon is that there is no fixed starting or ending point for the opened v-shaped trajectories while there exist an inflection point. Therefore, v-shape hand gestures performed by different participants may vary greatly due to their writing habits and thus these gestures are more easily to be confused. The results indicate that different templates and thresholds should be used for different participants and the templates and thresholds for each user should be collected and stored in a personal account during the system's first use. This means that a commercial application of the technology could potentially require a brief calibration period for a new user. Furthermore, labels 7 and 8 have high recognition accuracy mainly because circle trajectories are more distinctive than other gestures.

The participant recognition rate refers to the percentage of correctly identified gesture samples compared to the total number of samples for each participant. The statistical results of the participant recognition rates for both experiment types are shown in Fig. 11. It is evident that there is no significant difference between the recognition rate of different participants in the MTT experiment. Meanwhile, in the STT experiment, the recognition rates of participants 7 and 8 are fairly low compared to that of the participants, which can be attributed to the different inter-participant dexterity levels.

Moreover, the proposed method is compared with other methods which have similar gesture sets. For instance, Ji et al. [16] have proposed a hand gesture recognition algorithm with 3-axis accelerometer sensor. The preprocessed 3-axis waveforms from accelerometer are directly used as the input features and a fast dynamic time warping (FastDTW) is used as the recognizer. A fairly large gesture library has been proposed, but there are restricts on the rotation of hand gestures. However, the uncertainty of the dataset collected by a single inertial sensor may result in the decrease stability of the recognition rate for different hand gestures. Integration of multiple sensors for capturing different information of 3DHGs provide a good solution for this issue and this method is also adopted in present work. Moreover, there is no restraints on the rotation of hand gestures in present work that benefited from the comprehensive consideration of the rotation of the pen attitude.

Several previous studies also adopted multiple sensors to capturing the features in hand gestures, but the combination of different sensor may also affect the recognition of 3DHGs.

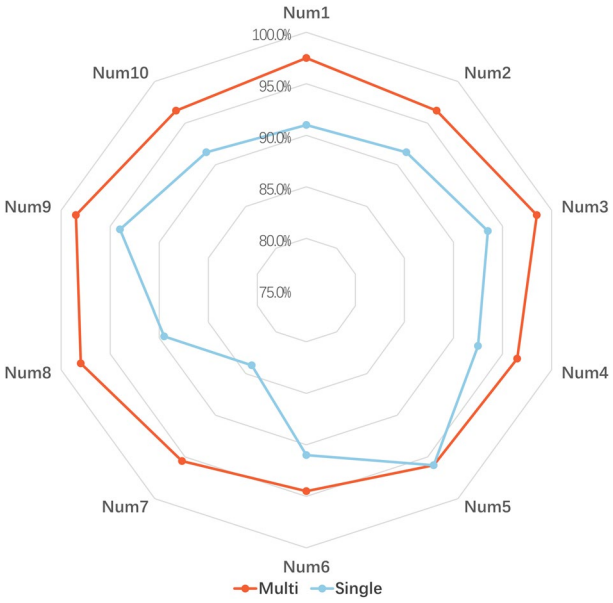


Fig. 11 Comparison of the experiment recognition rates in STT and MTT experiments

For example, the orientation sensor in Akan' s work is a virtual sensor which processes the data acquired from the remaining sensors, and the sampling rate of the orientation sensor will decrease slightly [2]. Moreover, the wearable and large-scale device may also decrease the flexibility and comfort of hand gesture interaction. The fusion of ultrasonic positioning and IMU data obtained by a pen-like device can overcome such drawback.

The comparison of the performance of proposed methodology and other state-of-the-art approaches in 3DHGs recognition is shown in Table 4. It seems obvious that the proposed 3DHGs recognition method with multiple templates and thresholds gives a higher recognition rate than previous work for the same hand gestures. Moreover, there is a better stability of recognition rate for 3DHGs of our proposed method. The rationality and effectiveness of the proposed recognition method for the 3DHGs are proved experimentally. The satisfactory performance of the method proposed in this work can be attributed

Table 4 Recognition rates of 3DHGs in different works and present work

Trajectory shape	Motional direction	Related work					Present work	
		[16]	[2]	[28]	[32]	[1]	MTT	STT
Line	Moving up	80%	-	-	89.2%	94%	98.5%	94%
	Moving down	100%	-	-	91.9%	93%	98.5%	94%
	Moving left	90%	-	96%	96.0%	95%	96%	91%
	Moving right	100%	-	84%	95.5%	95%	96%	92%
Circle	Clockwise	80%	66.6%	100%	-	-	99%	92.3%
	Anticlockwise	-	100%	-	-	-		
V-shape	Random	100%	-	-	-	-	94.5%	88.5%
Average		91.7%	83.3%	93.3%	93.3%	94.3%	97.1%	92.0%

to the comprehensive consideration of the attributes of 3DHGs and the multi-channel data are utilized in recognition process. Moreover, the MTT approach adopted in this work was specifically developed taking into consideration the different habits of participants, which is another crucial factor for the proposed method's high accuracy.

6 Conclusion

In this paper, a 3DHGs dataset for pen-based interaction is proposed based on the pen trajectory shape, motion direction, and pen attitude. All these attributes are considered in hand gestures recognition to give a more precision description of 3DHGs which can expands the system's interaction recognition capabilities remarkably. On this basis, a 3DHG primitive set was established. Then, a 3DHG recognition method based on ultrasonic positioning and inertial data is proposed, which involves first the identification of the trajectory shape, and then the motion direction and the pen body orientation according to a feature priority order for determining the hand gesture label.

Using the hardware of a 3D pen-based interactive device, a 3DHGs interaction scene and a data collection environment are established in Unity 3D software and further used to verify the effectiveness of the proposed method. STT and MTT experiment are performed according to the number of thresholds and template. The experimental results show that the 3DHGs recognition using the proposed method can achieve the accuracy of 99.5%. Multichannel data and multiple attributes show a great advantage in 3DHGs recognition which have not been studied before. The method proposed in this article allows the combination of specific interaction scenarios in daily life and the assignment of different interaction intentions to the same gestures, which has broad application prospects. However, the gestures library of present work is not copious enough and this will be improved continuously in future work. Moreover, the research work will focus on developing an improved algorithm for dynamic hand gesture recognition. In addition, the on-line learning or incremental learning to will be considered in our method for unknown hand gestures.

Acknowledgements This work was supported by the National Key Research and Development Program of China. (Grand 2016YFB1001301)

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

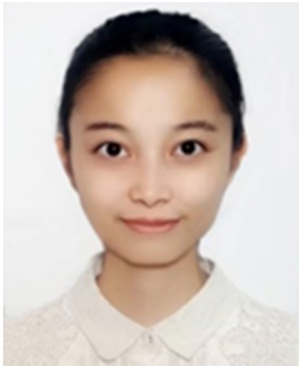
Competing Interests Human–computer interaction, 3D hand gesture recognition, 3D positioning, 3D pen-like interaction, Inertial sensor, Multichannel interaction.

References

1. Abid MR, Petriu EM, Amjadian E (2015) Dynamic Sign Language Recognition for Smart Home Interactive Application Using Stochastic Linear Formal Grammar. IEEE Transactions on Instrum Meas 64:596–605
2. Akan E, Tora H, Uslu B (2017) Hand gesture classification using inertial based sensors via a neural network. In: 2017 24th IEEE International Conference on Electronics, Circuits and Systems (ICECS). pp. 140–143

3. Akl A, Valaee S (2010) Accelerometer-based gesture recognition via dynamic-time warping, affinity propagation, & compressive sensing. In: 2010 IEEE International Conference on Acoustics, Speech and Signal Processing. pp. 2270–2273
4. Akl A, Feng C, Valaee S (2011) A Novel Accelerometer-Based Gesture Recognition System. IEEE Trans Sig Proc 59:6197–6205
5. Asano T, Honda S (2010) Visual interface system by character handwriting gestures in the air. In: 19th International Symposium in Robot and Human Interactive Communication. pp 56–61
6. Baddeley A (1992) Working memory. Sci 255:556–559
7. Che Y, Qi Y, Song Y (2019) Real-Time 3D Hand Gesture Based Mobile Interaction Interface. In: 2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct). pp 228–232
8. Chen M, AlRegib G, Juang B-H (2016) Air-Writing Recognition—Part I: Modeling and Recognition of Characters, Words, and Connecting Motions. IEEE Trans Human-Mach Syst 46:403–413
9. Chen J, Yu F, Yu J, Lin L (2020) A Three-Dimensional Ultrasonic Pen-Type Input Device With Millimeter-Level Accuracy for Human-Computer Interaction. IEEE Access 8:143837–143847
10. Chen J, Yu F, Yu J, Lin L (2021) A Three-Dimensional Pen-Like Ultrasonic Positioning System Based on Quasi-Spherical PVDF Ultrasonic Transmitter. IEEE Sens J 21:1756–1763
11. Cheng H, Yang L, Liu Z (2016) Survey on 3D Hand Gesture Recognition. IEEE Transactions on Circuits and Systems for Video Technology 26:1659–1673
12. Djemal A, Hellara H, Barioul R, et al (2022) Real-Time Model for Dynamic Hand Gestures Classification based on Inertial Sensor. In: 2022 IEEE 9th International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA). pp. 1–6
13. Dudak P, Sladek I, Dudak J, Sedivy S (2016) Application of inertial sensors for detecting movements of the human body. In: 2016 17th International Conference on Mechatronics - Mechatronika (ME). pp. 1–5
14. Hsieh C-C, Liou D-H (2015) Novel Haar features for real-time hand gesture recognition using SVM. J Real-Time Image Proc 10:357–370
15. Hsu Y-L, Chu C-L, Tsai Y-J, Wang J-S (2014) An inertial pen with dynamic time warping recognizer for handwriting and gesture recognition. IEEE Sens J 15:154–163
16. Ji Z, Li Z-Y, Li P, An M (2015) A new effective wearable hand gesture recognition algorithm with 3-axis accelerometer. In: 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD). pp. 1243–1247
17. Katsura S, Ohishi K (2007) Acquisition and Analysis of Finger Motions by Skill Preservation System. IEEE Transactions on Industrial Electronics 54:3353–3361
18. Keskin C, Erkan AŞ, Akarun L (2003) REAL TIME HAND TRACKING AND 3D GESTURE RECOGNITION FOR INTERACTIVE INTERFACES USING HMM
19. Long AC, Landay JA, Rowe LA (1997) PDA and Gesture Use in Practice: Insights for Designers of Pen-based User Interfaces
20. Marasovic T, Papic V, Zanchi V (2015) LMNN metric learning and fuzzy nearest neighbour classifier for hand gesture recognition. J Multimod User Interfac 9:211–221
21. Mingji H, Lin L (2016) Study on brush gesture feature recognition based on the 3D point cloud. In: 2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA). pp 1634–1638
22. Pan T-Y, Kuo C-H, Liu H-T, Hu M-C (2019) Handwriting Trajectory Reconstruction Using Low-Cost IMU. IEEE Trans Emerg Topics Comput Intel 3:261–270
23. Pezzuoli F, Corona D, Corradini ML (2020) Recognition and Classification of Dynamic Hand Gestures by a Wearable Data-Glove. SN Comput Sci 2:5
24. Pomboza-Junez G, Terriza JH (2016) Hand gesture recognition based on sEMG signals using Support Vector Machines. In: 2016 IEEE 6th International Conference on Consumer Electronics - Berlin (ICCE-Berlin). pp 174–178
25. Qi J, Jiang G, Li G et al (2019) Intelligent Human-Computer Interaction Based on Surface EMG Gesture Recognition. IEEE Access 7:61378–61387
26. Sali Shajideen SM, Preetha VH (2018) Human-Computer Interaction System Using 2D and 3D Hand Gestures. In: 2018 International Conference on Emerging Trends and Innovations In Engineering And Technological Research (ICETIETR). pp 1–4
27. Takahashi H, Kitazono Y (2016) Integration of Hand Gesture and Multi Touch Gesture with Glove Type Device. In: 2016 4th Intl Conf on Applied Computing and Information Technology/3rd Intl Conf on Computational Science/Intelligence and Applied Informatics/1st Intl Conf on Big Data, Cloud Computing, Data Science & Engineering (ACIT-CSII-BCD). pp 81–86
28. Teachasrisaksakul K, Wu L, Yang G-Z, Lo B (2018) Hand Gesture Recognition with Inertial Sensors. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 3517–3520

29. Wang J-S, Chuang F-C (2012) An Accelerometer-Based Digital Pen With a Trajectory Recognition Algorithm for Handwritten Digit and Gesture Recognition. IEEE Trans Indus Electron 59:2998–3007
30. Xin Y, Bi X, Ren X (2011) Acquiring and Pointing: An Empirical Study of Pen-Tilt-Based Interaction. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, pp 849–858
31. Xu R, Zhou S, Li WJ (2012) MEMS Accelerometer Based Nonspecific-User Hand Gesture Recognition. IEEE Sensors Journal 12:1166–1173
32. Zhu Y, Yuan B (2014) Real-time hand gesture recognition with Kinect for playing racing video games. In: 2014 International Joint Conference on Neural Networks (IJCNN). pp 3240–3246
33. Zhu C, Yang J, Shao Z, Liu C (2021) Vision Based Hand Gesture Recognition Using 3D Shape Context. IEEE/CAA J Automatica Sinica 8:1600–1613



Siyu Liu received her BS degree in Communication Engineering, the MS degree in Electronics and Communication Engineering from the College of Communication engineering of Jilin University, Changchun, China, in 2019 and 2022 respectively. Her research interests include inertial sensing, hand gesture recognition and machine learning.

Jian Chen received a BS degree in Electrical and Information Engineering, an MS degree in Signal and Information Processing, and a Ph.D. degree in Communication and Information Systems from Jilin University, Changchun, China, in 2001, 2004, and 2007, respectively. He is currently a professor at the College of Communication Engineering, Jilin University. From 2014 to 2015, he was a visiting scholar at the School of Electronic and Electrical Engineering of the University of Leeds. His research interests include ultrasonic positioning systems and array signal processing.

Cheng Wang received the BS degree in Microelectronics Technology from the University of Electronics Science and Technology of China, Chengdu, China, in 2015. In 2019, he received his MS degree in communication engineering from Jilin University, Changchun, China. His research interests include 3D human-computer interaction, computer graphics and computational photography, computer vision, and virtual reality.

Lin Lin received her BS degree in Measurement and Control Technology and Instrumentation, the MS and Ph.D. degrees in Communication and Information Systems from Jilin University, Changchun, China, in 2001, 2004, and 2007, respectively. From 2007 to 2019, she was a lecturer and is currently an associate professor at the College of Communication Engineering, Jilin University. Her research interests include ultrasonic positioning systems and digital signal processing applications.