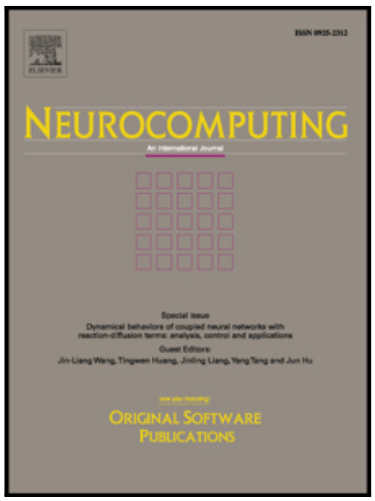

3D human gesture capturing and recognition by the IMMU-based data glove

Bin Fang , Fuchun Sun , Huaping Liu , Chunfang Liu

PII: S0925-2312(17)31405-4
DOI: 10.1016/j.neucom.2017.02.101
Reference: NEUCOM 18795

To appear in: *Neurocomputing*

Received date: 18 September 2016
Revised date: 5 February 2017
Accepted date: 11 February 2017

Please cite this article as: Bin Fang , Fuchun Sun , Huaping Liu , Chunfang Liu , 3D human gesture capturing and recognition by the IMMU-based data glove, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2017.02.101

3D human gesture capturing and recognition by the IMMU-based data glove

Bin Fang*, Fuchun Sun, Huaping Liu, Chunfang Liu

Tsinghua National Laboratory for Information Science and Technology,Department of Computer Science and Technology, Tsinghua University, Tsinghua National Laboratory for Information Science and Technology, Beijing, China
bravebin@mail.tsinghua.edu.cn

Abstract. Gestures recognition provides an intelligent, natural, and convenient way for human–robot interaction (HRI). This paper presents a novel data glove for gestures capturing and recognition based on inertial and magnetic measurement units (IMMUs), which are made up of three-axis gyroscopes, three-axis accelerometers and three-axis magnetometers. The proposed data glove has eighteen low-cost IMMUs, which are compact and small enough to wear. The gestures included the three-dimensional motions of arm, palm and fingers are completely captured by the data glove. Meanwhile, we attempt to use extreme learning machine (ELM) for gesture recognition which has not found yet in the relevant application. The ELM-based recognition methods **for both** static gestures and dynamic gestures are respectively presented. The experimental results of gestures capturing and recognition verify the effectiveness of the proposed methods.

Keywords: gestures recognition, capture, inertial and magnetic measurement unit, extreme learning machine

1 Introduction

Gestures are expressive, meaningful body motions involving physical movements of the fingers, hands, arms with the intent to convey meaningful information or to communicate with the environment. With the rapid development of computer technology, various approaches of human computer interaction have been proposed in these years. Human-computer interaction with hand gestures plays a significant role in these modalities. Therefore, hand-gesture-based methods stand out from other approaches by providing a natural way of interaction and communication.

Recently, various gesture capturing and recognition technologies have been proposed. These studies can be divided into two categories, based on their motion capture mechanism: vision-based or glove-based [1]. Vision-based techniques rely on image processing algorithms to extract motion trajectory and posture information. On the other hand, glove-based techniques rely on physical interaction with the user. Based on vision gesture method, users generally do not need to wear collection equipment

and move freely, but it is easily affect by illumination, occlusion, placement of camera and other environmental factors [2]. In comparison, glove-based techniques are easy to implement, and generally provide more reliable motion data [3]. Different types of sensory gloves have been developed, both commercial and prototype ones. The commercial products [4] usually use expensive motion-sensing fibers and resistive-bend sensors, and are consequently too costly for the consumer market [5]. Consequently, the prototype data gloves are developed to lower the cost of such equipment [6]. The flex sensors or bending sensors are integrated into the data gloves. However, the above sensors just measure the relative orientation of articulated segments by mounting the sensor across the joint of interest. This requires an accurate alignment of sensors with particular joint. Moreover, re-calibration during utilization is necessary to mitigate estimation errors due to sensor displacements. A general disadvantage of data gloves is the lack of user customization for individual subject' hands and obstruction of tactile sensing from the palmar surface of the hand. Often this inherently goes with mounting space required for embedding the sensors in clothing. To overcome the shortcomings, the inertial and magnetic sensors are induced.

In recent years, with the development of MEMS technology, micro inertial devices have numerous advantages such as small size, low power consumption, large dynamic range, low cost and so on. It has gradually become the main sensor of human motion capture [7]. Meanwhile, the magnetic sensors are used together with inertial sensors for accurate and drift free orientation estimation [8]. Inertial and magnetic measurement unit (IMMU) has proven to be an accurate approach in estimating body segment orientations without the external actuators or cameras [9]. It is non-obtrusive, comparably cost effective, and easy to integrate. It demonstrates higher correlation and lower error compared with a research-used visual motion capture system when the same motions are recorded [10]. The low-cost, small wearable inertial and magnetic sensors are becoming increasingly popular for data glove. The KHU-l data glove [11] is developed by using six three-axis accelerometers, but it only captures the recognition of several kinds of gestures. In [12] the data glove is developed based on sixteen micro inertia sensors, which can capture the movements of each finger and palm, but the information of the heading angle is missing. In [13] the inertial and magnetic measurement unit is used, but it only uses the four inertial and magnetic measurement units, it is unable to obtain the information of each finger joint. Power-Glove [14] is developed that include six nine-axis micro inertial sensors and ten six-axis micro inertial sensors. It covers each joint of the palm and fingers, and motion characteristics can be better evaluated. However, it is not fully using nine micro inertial sensors, in some state heading angle solution is instability that would lead to the estimation errors of the joint angle. The research shows that the current gesture capture device does not take into account the three-dimensional motion of the arm and hand. Therefore, we use the inertial and magnetic measurement units to develop a novel data glove, which can fully capture the gesture motion information of the forearm, upper arm, palm and fingers.

On the other hand, various gesture recognition techniques such as k-Nearest Neighbor(k-NN) [15], ANNs (artificial neural networks) [16], and SVM (Support Vector Machine) [17] has (have) been applied to the field of hand gesture recognition.

Nonetheless, it is known that both them face some challenging issues as follows:(1) slow training speed (2) trivial human intervene (3) large computational quantity (4) poor generalization ability [18]. However, compared with those machine learning algorithms, extreme learning machine (ELM) has better generalization performance at a much faster learning speed [19-20]. In addition, ELM is insensitive to the parameters [21-22]. According to our investigation, ELM has not been applied in the field of gesture recognition based on the data glove. Hence, we use ELM to recognize the gestures that are captured by the proposed data glove.

The remainder of this paper is organized as follows: In Section 2 the gesture capturing method is introduced. Section 3 presents the theories of ELM-based gesture recognition. Section 4 gives the experimental results; while Section 5 concludes this paper.

2 Gesture Capturing Method

In this section, the gestures capture methods are presented. First, the design of the inertial and magnetic measurement unit is described, and the data glove is presented. Then the orientations filter based quaternion-based extended Kalman filter is deduced, the gesture is determined by the 3D orientations of the each IMMU.

2.1 Data glove design

Firstly the inertial and magnetic measurement units are designed. Commercial IMMU, such as Xsens-MTi [23] or Shimmer [24] are not small enough to be placed precisely and stably on the key points of the fingers or arms. Here, the MPU9250 [25] that combines 9-axis inertial and magnetic sensors in a very small package is used. The MPU9250 sensor is mounted on the solid PCB with the dimensions of 10×15×2.6mm and a weight of about 6g. Hence the low-cost, low-power and light-weight IMMU is developed. The small IMMU can be fastened to the glove, which makes the glove more appealing and easier to be used. Some of the important physical properties of the sensor board are compared with Xsens-MTi and Shimmer in Table 1. Then the connectivity of the IMMUs are designed. Different types of communication architectures are described in [26]. Wireless networking approaches for connections are potable, but the complexity and energy consumption of wireless networks is increased. Hence a wired approach is used in [27], all sensor units are directly connected to a central controlling unit using cables. These results in a very complex wiring. In the present work, a cascaded wiring approach is proposed and developed by exploiting the master SPI bus of each IMMU. This approach simplifies wiring without any need for extra components. Since data reading from a string of IMMUs, the MCU does not switch to all the IMMUs to fetch the data, which results in a lower power consumption. In order to increase the flexibility, textile cables are used to connect the IMMU to each other and to the MCU. Here the STM32F4 microcontroller is used to develop the MCU.

Table 1. Comparison of the proposed IMMU with other products

IMMU	Xsens-MTi	Shimmer	Proposed unit
Price[¥]	19900	2190	85
Dimension[mm]	38×5×21	52×32×17	10×15×2.6
Weight[g]	30	25	6
Power consumption[mW]	350	300	50

After determining the above designs, the wear design of data glove can be determined. There are eighteen segments of the arm and hand, hence eighteen IMMUs are used to cover all of the segments. Six strings are designed that each string deploys three IMMUs for corresponding segments. The IMMUs' data is sampled and computed by the MCU and subsequently transmitted via Bluetooth to the external devices. The proposed data glove is shown in Figure 1. The proposed data glove is designed based on the low-cost IMMU, which can capture the more information of motion than the traditional sensors. The traditional sensors of data glove like fiber or hall-effect sensor are frail. Nevertheless, the board of inertial and magnetic sensor that is an independent unit. It is more compact, more durable and more robust. Commercial data gloves are too costly for the consumer market, but the proposed data glove in the paper is low-cost. Moreover, the proposed data glove can not only capture the motion of hand but also capture the motion of arm, and the estimated results of motion are outputting real time.

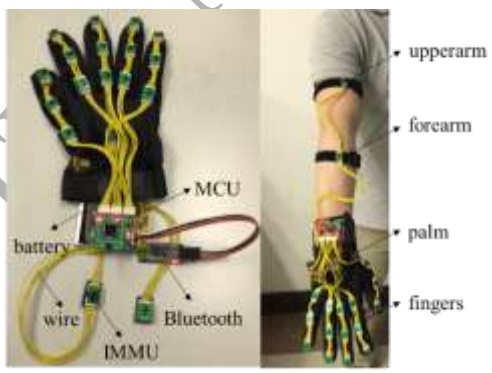

Fig. 1. The proposed data glove and wearing demonstration

2.2 Proposed methods for gesture capturing

The orientations estimation algorithm is proposed in the section. The frames are shown in Figure 2. In this figure, global frame is N and local reference frames are A_1 (upper-arm), A_2 (forearm), H (hand), P_i (proximal), M_i (medial), D_i (distal), where $i = 1,2,3,4,5$, which represent the Thumb, the Index, the Middle, the Ring, and the

Little, respectively. The global reference frame z-axis is defined along the axial axis (from the head to the feet) of the subject, the y-axis along the sagittal (from the left shoulder to the right shoulder) axis and the x -axis along the coronal axis (from the back to the chest). The local frames are parallel to the global frame. Meanwhile, two assumptions about data glove in use are made: (1) The body keeps static, just arm and hand are in motion; (2) The local static magnetic field is homogeneous throughout the whole arm.

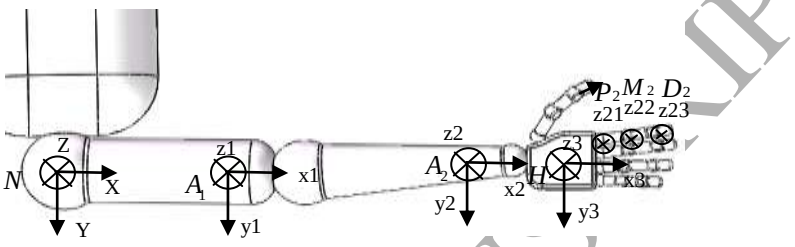

Fig. 2. The frames of the data glove

Based on the measured 3D angular velocity, acceleration and magnetic field of one single IMMU, it is possible to stably estimate its orientation with respect to a global coordinate system. The global coordinate N and each coordinate B of IMMUs are shown in Fig.2. The transformation between the representations, relative to N and B , of a 3×1column-vector $\boldsymbol{x}$ is expressed as

$$\boldsymbol{x}^b = \boldsymbol{C}_n^b[\boldsymbol{q}]\boldsymbol{x}^n \tag{1}$$

where quaternion $\boldsymbol{q} = [q_0; \boldsymbol{Q}]$, $[\boldsymbol{Q}]$ is the anti-symmetric matrix given by

$$[\boldsymbol{Q}] = \begin{bmatrix} 0 & Q_3 & -Q_2 \\ -Q_3 & 0 & Q_1 \\ Q_2 & -Q_1 & 0 \end{bmatrix}.$$

The attitude matrix $\boldsymbol{C}$ is related to the quaternion by

$$\boldsymbol{C}(q) = \left(q_0^2 - \boldsymbol{Q}\cdot\boldsymbol{Q}\right)\boldsymbol{I} + 2\boldsymbol{Q}\boldsymbol{Q}^T + 2q_0[\boldsymbol{Q}] \tag{2}$$

where, $\boldsymbol{I}$ is the identity matrix.

The gyro, the accelerometer and the magnetometer are sensor triplets with perpendicular sensitivity axes. Their output in response to the angular velocity, total acceleration (gravity and acceleration), earth's magnetic field are expressed, respectively, by

$$\boldsymbol{\omega} = \boldsymbol{\omega}_t + \boldsymbol{v}_g \tag{3}$$

$$\boldsymbol{a} = \boldsymbol{C}_n^b(q)\boldsymbol{g} + \boldsymbol{v}_a \tag{4}$$

$$\boldsymbol{m} = \boldsymbol{C}_n^b(q)\boldsymbol{h} + \boldsymbol{v}_m \tag{5}$$

where $\boldsymbol{v}_g$, $\boldsymbol{v}_a$, $\boldsymbol{v}_m$ are assumed uncorrelated white Gaussian measurement noise, with null mean and covariance matrix $\mathrm{K}_g = \sigma_g^2\boldsymbol{I}$, $\mathrm{K}_a = \sigma_a^2\boldsymbol{I}$, $\mathrm{K}_m = \sigma_m^2\boldsymbol{I}$

The state vector is composed of the rotation quaternion. The state transition vector equation is,

$$\boldsymbol{x}_{k+1} = q_{k+1} = \Phi\left(T_s, \omega_k\right) + w_k = \exp\left(\Omega_k T_s\right) q_k + {}^{q}w_k \tag{6}$$

$${}^{q}w_k = -\frac{T_s}{2}\Gamma_k\, {}^{q}\boldsymbol{v}_k = -\frac{T_s}{2}\begin{bmatrix}\left[e_k \times\right] + q_{4k}\boldsymbol{I} \\ -e_k^T\end{bmatrix} {}^{g}\boldsymbol{v}_k \tag{7}$$

where the gyro measurement noise vector ${}^{g}\boldsymbol{v}_k$ is assumed small enough that a first order approximation of the noisy transition matrix is possible.

Then the process noise covariance matrix Q_k will have the following expression,

$$Q_k = \left(T_s/2\right)^2 \Gamma_k \Gamma_g \Gamma_k^T \tag{8}$$

The measurement model is constructed by stacking the accelerometer and magnetometer measurement vectors,

$$\boldsymbol{z}_{k+1} = \begin{bmatrix}\boldsymbol{a}_{k+1} \\ \boldsymbol{m}_{k+1}\end{bmatrix} = \begin{bmatrix}\boldsymbol{C}_n^b\left(q_{k+1}\right) & 0 \\ 0 & \boldsymbol{C}_n^b\left(q_{k+1}\right)\end{bmatrix}\begin{bmatrix}\boldsymbol{g} \\ \boldsymbol{h}\end{bmatrix} + \begin{bmatrix}{}^{a}\boldsymbol{v}_{k+1} \\ {}^{m}\boldsymbol{v}_{k+1}\end{bmatrix} \tag{9}$$

The covariance matrix of the measurement model $\boldsymbol{R}_{k+1}$ is,

$$\boldsymbol{R}_{k+1} = \begin{bmatrix}{}^{a}\boldsymbol{R}_{k+1} & 0 \\ 0 & {}^{m}\boldsymbol{R}_{k+1}\end{bmatrix} \tag{10}$$

where the accelerometer and magnetometer measurement noise ${}^{a}\boldsymbol{v}_{k+1}$ and ${}^{m}\boldsymbol{v}_{k+1}$ are uncorrelated zero-mean white noise process, the covariance matrices of which are ${}^{a}\boldsymbol{R}_{k+1} = \sigma_a^2\boldsymbol{I}$ and ${}^{m}\boldsymbol{R}_{k+1} = \sigma_m^2\boldsymbol{I}$, respectively.

Because of the nonlinear nature of (9), the EKF approach requires that a first-order Taylor-Mac Laurin expansion is carried out around the current state estimation by computing the Jacobian matrix,

$$\boldsymbol{H}_{k+1} = \frac{\partial}{\partial \boldsymbol{x}_{k+1}}\boldsymbol{z}_{k+1}\bigg|_{\boldsymbol{x}_{k+1}=\boldsymbol{x}_{k+1}^-} \tag{11}$$

Then the orientations are estimated by the following EKF equations,

Compute the *a priori* state estimate

$$\boldsymbol{x}_{k+1}^- = \Phi\left(\boldsymbol{T}_s, \boldsymbol{\omega}_k\right)\boldsymbol{x}_k \tag{12}$$

Compute the a priori error covariance matrix

$$\boldsymbol{P}_{k+1}^- = \Phi\left(\boldsymbol{T}_s, \boldsymbol{\omega}_k\right)\boldsymbol{P}_k\Phi\left(\boldsymbol{T}_s, \boldsymbol{\omega}_k\right)^T + \boldsymbol{Q}_k \tag{13}$$

Compute the Kalman gain

$$\boldsymbol{K}_{k+1} = \boldsymbol{P}_{k+1}^-\boldsymbol{H}_{k+1}^T\left(\boldsymbol{H}_{k+1}\boldsymbol{P}_{k+1}^-\boldsymbol{H}_{k+1}^T + \boldsymbol{R}_{k+1}\right)^{-1} \tag{14}$$

Compute the *a posteriori* state estimate

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_{k+1}^{-} + \boldsymbol{K}_{k+1}\left[\boldsymbol{z}_{k+1} - f\left(\boldsymbol{x}_{k+1}^{-}\right)\right] \tag{15}$$

Compute the *a posteriori* error covariance matrix

$$\boldsymbol{P}_{k+1} = \boldsymbol{P}_{k+1}^{-} - \boldsymbol{K}_{k+1}\boldsymbol{H}_{k+1}\boldsymbol{P}_{k+1}^{-} \tag{16}$$

According to the above algorithm, the three orientations of each IMMU can be estimated. Hence, the 3D gestures integrated by the orientations of upperarm, forearm, palm and fingers are determined. Then, the orientations of eighteen IMMUs are directly used to recognize the gestures.

3 ELM-based Gestures Recognition

In this section, the ELM-based gesture recognition method is proposed. The framework of the gesture recognition using ELM is described. And the ELM-based recognition methods of static gestures and dynamic gestures are respectively presented.

3.1 Architecture of the gesture recognition

The framework of the gestures recognition based on the proposed data glove using ELM can be divided into three stages. The first stage is establishing a gestures database. Firstly, the motion data of different gestures is collected through the proposed data glove. Then, we establish a gestures dataset which includes static gestures and dynamic gestures. The second stage is training classifier. On the basis of the gesture dataset, we extract the 54-Dimension hand feature of each gesture and express it. The third stage is experiment and analysis. We collect the motion data from different participants of various gestures, and use the trained classifiers for gesture recognition. The architecture of the system is shown in Figure 3.

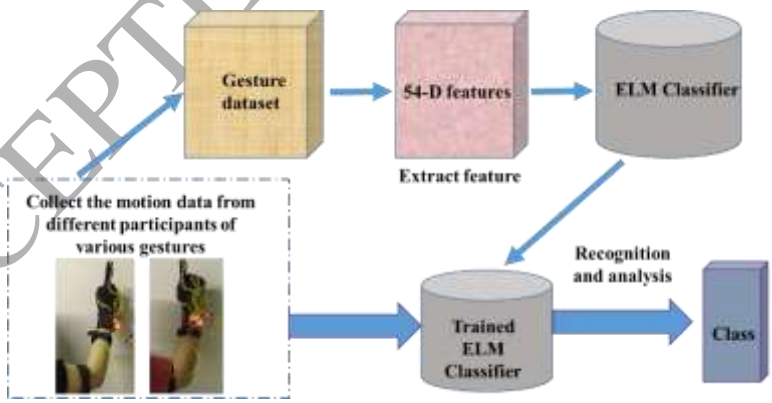

Fig. 3. Architecture of the gestures recognition based on data glove using ELM.

3.2 ELM-based static gestures recognition

ELM was first proposed by Huang et al., which randomly generates the input weights and hidden layer biases of SLFNs and then determines the output weights analytically. And ELM can be biologically inspired, provide significant advantages such as fast learning speed, independent from implementation, and minimal human intervention [28]. The model of ELM is shown in Figure 4.

If the input data is $\boldsymbol{x}$, then we can obtain the output function of the L hidden layer node as,

$$\begin{cases} f_L(\boldsymbol{x}) = \sum_{i=1}^{L} \beta_i g_i(\boldsymbol{x}) = \sum_{i=1}^{L} \beta_i G_i(\boldsymbol{w}_i, \boldsymbol{b}_i, \boldsymbol{x}) \\ w_i \in C^d, x_i \in C^d, \beta_i \in C. \end{cases} \tag{17}$$

where β_i is the output weight vector of the node of the i_{th} hidden layer, $g_i(x)$ denotes hidden nodes nonlinear piecewise continuous activation functions.

According to the theory of Bartlett [29], the method based on the least weight is used to calculate the output weights, and the ELM can obtain the minimum error solution through the minimum norm, which will achieve a great general performance. Given N training samples $(\boldsymbol{x}_i, \boldsymbol{t}_i)$, the output of L hidden layer nodes is,

$$f(\boldsymbol{x}) = \sum_{i=1}^{L} \beta_i \boldsymbol{G}(\boldsymbol{w}_i, \boldsymbol{b}_i, \boldsymbol{x}) = \beta \cdot h(\boldsymbol{x}) \tag{18}$$

where $h(x)$ is the output vector of the hidden layer, the parameters of the hidden layer nodes are randomly assigned, βi is the weight vector connecting the i-th hidden neuron and output neurons.

Thus, the matrix expression of this linear system is,

$$\boldsymbol{H} \cdot \beta = \boldsymbol{T} \tag{19}$$

$$\boldsymbol{H} = \begin{pmatrix} \boldsymbol{G}(\boldsymbol{w}_1, \boldsymbol{b}_1, \boldsymbol{x}_1) & \cdots & \boldsymbol{G}(\boldsymbol{w}_L, \boldsymbol{b}_L, \boldsymbol{x}_1) \\ \vdots & \cdots & \vdots \\ \boldsymbol{G}(\boldsymbol{w}_1, \boldsymbol{c}_1, \boldsymbol{x}_N) & \cdots & \boldsymbol{G}(\boldsymbol{w}_L, \boldsymbol{b}_L, \boldsymbol{x}_N) \end{pmatrix}_{N \times L} \tag{20}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times d}, \quad \boldsymbol{T} = \begin{bmatrix} \boldsymbol{t}_1^T \\ \vdots \\ \boldsymbol{t}_L^T \end{bmatrix}_{N \times d} \tag{21}$$

Based on the input $\boldsymbol{x}_i$, the network matrix $\boldsymbol{H}$ is outputs of hidden layers, in which the i-th line represent the output vector of hidden layer. According to all input $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$, the i-th column represents the output of the i-th hidden layer neuron. The minimum norm the least-square solution to the linear system is:

$$\left|\boldsymbol{H}\cdot\hat{\beta}-\boldsymbol{T}\right|=\min_{\beta}\left|\boldsymbol{H}\cdot\beta-\boldsymbol{T}\right| \tag{22}$$

Therefore,

$$\hat{\beta}=\boldsymbol{H}^{\dagger}\boldsymbol{T} \tag{23}$$

where $\boldsymbol{H}^{\dagger}$ is the Moore-Penrose generalized inverse of matrix $\boldsymbol{H}$.

In order to improve the ELM with better generalization capabilities in comparison with the least square solution-based ELM, which requires randomly generated input weights, Huang et al. proposed the use of kernel methods in the design of ELM and suggested adding a positive value $1/C$ (where C is a user-defined parameter) for the calculation of the output weights such that:

$$\beta=\boldsymbol{H}^{T}\left(\frac{\boldsymbol{I}}{\boldsymbol{C}}+\boldsymbol{H}\boldsymbol{H}^{T}\right)^{-1}\boldsymbol{T} \tag{24}$$

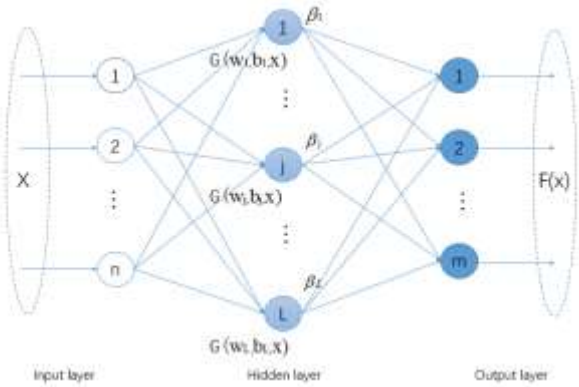

Fig. 4. The model of the ELM

Kernel-based ELM can be represented as follows:

$$\begin{aligned}\boldsymbol{K}_{ELM}(\boldsymbol{x}_i,\boldsymbol{x}_j)&=\boldsymbol{h}(\boldsymbol{x}_i)\cdot\boldsymbol{h}(\boldsymbol{x}_j)\\&=[\boldsymbol{G}(\boldsymbol{w}_1,\boldsymbol{b}_1,\boldsymbol{x}_i),\ldots,\boldsymbol{G}(\boldsymbol{w}_L,\boldsymbol{b}_L,\boldsymbol{x}_i)]^{T}\cdot[\boldsymbol{G}(\boldsymbol{w}_1,\boldsymbol{b}_1,\boldsymbol{x}_j),\ldots,\boldsymbol{G}(\boldsymbol{w}_L,\boldsymbol{b}_L,\boldsymbol{x}_j)]^{T}\end{aligned} \tag{25}$$

Because of the parameter of $(\boldsymbol{w},\boldsymbol{b})$ is randomly assigned, the $h(\cdot)$ is also randomly generated, so the dual kernel optimization function is:

$$\begin{aligned}&\text{minimize}:\mathbf{L}_D=\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}t_i t_j\boldsymbol{K}_{ELM}(\boldsymbol{x}_i,\boldsymbol{x}_j)\alpha_i\alpha_j-\sum_{i=1}^{N}\alpha_i\\&\text{subject to}:0\le\alpha_i\le C,i=1,\ldots,N\end{aligned} \tag{26}$$

which can be seen from the above, the kernel ELM is a combination of the optimization method of nuclear learning and standard to find the optimal solution. Due to the

relatively weak optimization constraints, the kernel ELM has better generalization ability.

3.3 ELM-based dynamic gesture recognition

In order to cope with the dynamic gestures that are time series, a dynamic time warping (DTW) algorithm is used. DTW is a nonlinear warping technique combined with time warping and distance calculating, which can complete the matching process of the model with global or local extension, compression or deformation. DTW algorithm is to establish a scientific time alignment matching path between the characteristics of the test pattern and the reference pattern.

Suppose we have two time series T_i and R_j : the test pattern feature vector sequence is,

$$T = (t_1, t_{2,} \cdots\cdots, t_i) \tag{27}$$

And the reference pattern feature vector sequence is,

$$R = (r_1, r_2, \cdots\cdots, r_j) \tag{28}$$

where *i, j* are the time serial number. A warping path W defines a mapping between T_i and R_j , so we have

$$W = \{w_1, w_2, \cdots\cdots, w_N\} \tag{29}$$

where *N* is the last warping path, $w_n = (i, j)_n$ is the *n*-th mapping between the *i*-th of test pattern feature vector and the *j*-th of reference pattern feature vector.

So the minimizing warping cost as follows,

$$DTW(T, R) = \min \sqrt{\sum_{n=1}^{N} w_n} \tag{30}$$

This path can be found using dynamic programming to evaluate the following recurrence which defines the cumulative distance $l(T_i, R_j)$ and the current cell distance $d(T_i, R_j)$,

$$l(T_i, R_j) = d(T_i, R_j) + \min\{l(T_{i-1}, R_{j-1}), l(T_{i-1}, R_j), l(T_i, R_{j-1})\} \tag{31}$$

Currently, the popular kernel functions of ELM can be as follows：

(1) polynomial-kernel: $K(x, x_i) = [a(x, x_i) + c]^q$;

(2) RBF-kernel: $K(x, xi) = \exp\left(-\gamma \left\| x - x_i \right\|^2\right)$;

(3) sigmoid-kernel: $K(x, xi) = \tan h\left[a(x, x_i) + c\right]$.

In this paper, we attempt to combine DTW and RBF-kernel to achieve the aim that the dynamic gesture recognition based on time series. It can be expressed as

$$K(R,T) = \exp[-\gamma DTW(R,T)^2] \quad (32)$$

where γ is a prescribed adjusting parameter.

4 Experiments and Results

The experiments concentrate on the contribution of this paper, i.e. the development of a low-cost data glove for gesture capture, and the ELM-based gesture recognition. Hence, the proposed gestures tracking and recognition algorithms will be evaluated in the following experiments.

4.1 Motion capture experiments and results

The eighteen IMMUs are used to capture the arm and hand motion. Sixteen IMMUs that are on the glove wear in the hand. The rest of two IMMUs are bound in upper-arm and forearm by the magic sticker, respectively. The MCU board and lithium battery are worn on the wrist. The battery is under the MCU board. Its output voltage is 5v, and it continuously supplies power for the IMMUs and MCU about three hours. The MCU processes the raw data and estimates results, encapsulates them into a packet, and then sends the packet to the PC by Bluetooth. The baud rate for transmitting data is 115200 bps. The frequency can be reached 50 Hz. By using this design, the motion capture can be demonstrated by the virtual model on the PC immediately. The system architecture is shown as Figure 5.

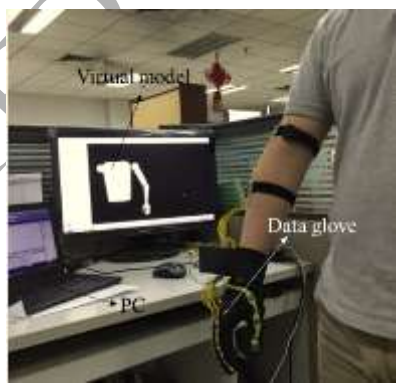

Fig. 5. The first captured gesture

To verify the effective of orientation estimation algorithm, three orientations of the IMMUs are evaluated by the different situations. The results of six IMMUs, which are respectively on upper-arm, forearm, palm, proximal of the forefinger, medial of the forefinger, distal of the forefinger, are shown in Fig.6. In 0s~4s, the arm and hand keep static. In 4s~12s, only the upper-arm is swing up and down. In 18s~25s, only the

forearm is swing right and left. In 30s~39s, only the palm is swing up and down. In 43s~67s, only the fingers are bending and opening. As shown in Fig.5, the orientations of the arm, palm and fingers can be determined in different situations, moreover, the movements can also be easily distinguished. Meanwhile, the accuracy of the results is assessed by the statistics. The root mean square error (RMSE) of the orientations are less than 0.5°. The detail results are listed in Table 2.

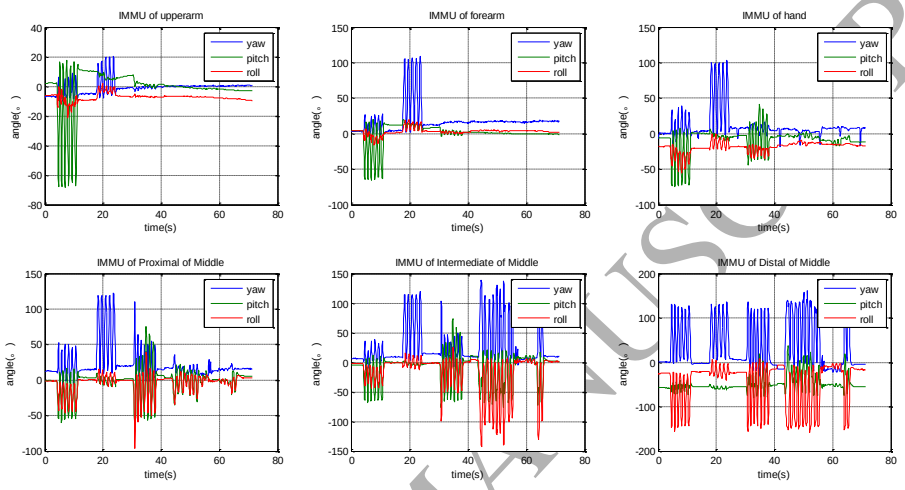

Fig. 6. 3D Orientations of the IMMUs

Table 2. RMSE of the orientations

RMSE(°)	Yaw	Pitch	roll
value	0.44	0.22	0.31

4.2 Gesture recognition experiments and results

The experiments of the static gestures and dynamic gestures are respectively implemented to verify the effectiveness of the ELM-based recognition method. The ten static gestures and sixteen dynamic gestures are captured for identification in the following experiments.

4.2.1 Static gestures recognition

As a basis for the experiments, we recorded a dataset containing 10 different classes of static gestures, which are the numbers from ‘one’ to ‘ten’. The gestures are shown in Figure 7. The data have been recorded with 2 participants. The 54-dimention orientations of the data glove are used to express the gestures. Hence we use the 54-dimention feature for gesture recognition classification. Table 3 summarizes that compared with other methods, which shows that ELM-kernel can achieve bet-

ter performance. Moreover, we can see from Table 3 that the time it takes to ELM is far less than SVM. The average gesture recognition classification accuracy of ELM-Kernel is higher than that of ELM and SVM. Figure 8 shows the confusion matrix across all 10 classes. Most model confusions are evidently showing that the ELM kernel method has a better gesture recognition accuracy compared with the ELM and SVM.

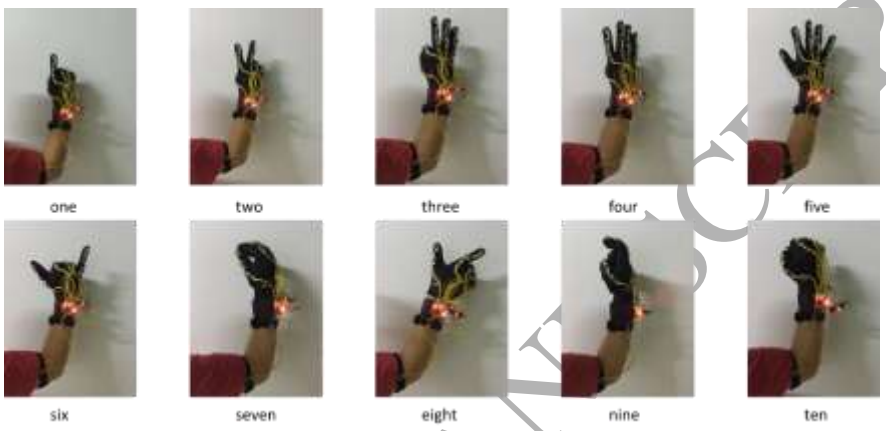

Fig. 7. The static gestures of numbers

Table 3. Gesture recognition accuracy and train time for different approaches

	ELM	**ELM-kernel**	**SVM**
accuracy	68.05%	89.59%	83.65%
Train time (s)	1.6094	12.5456	1886.5

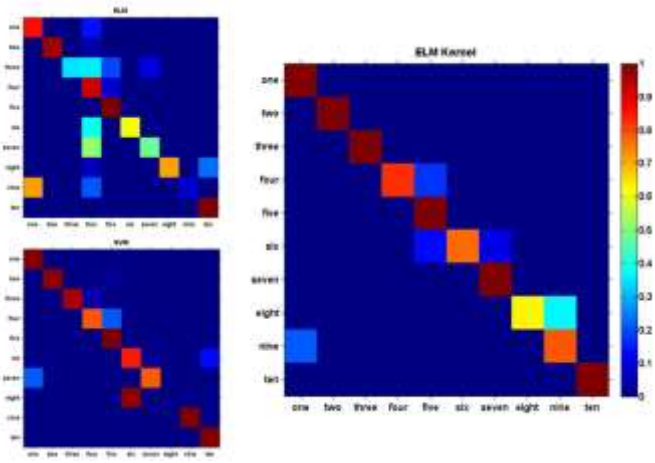

Fig. 8. The confusion matrix of ELM, SVM, ELM Kernel

4.2.2 Dynamic gestures recognition

In the dynamic gestures experiments, we recorded a dataset containing 16 different classes of dynamic gestures, the directions of which included up, down, left, right, up-left, up-right, down-left, down-right. And the eight gestures are waved only by hand. The gestures are shown in Figure 9. The data have also been recorded with 2 participants. We use the DTW distance feature for gesture recognition classification. The ELM-DTW based method and SVM-DTW based method are implemented to compare the performance of dynamic gesture recognition. Table 4 summarizes the results of recognition accuracy, and ELM-DTW based gesture recognition method achieves better performance. Figure 10 and Figure 11 respectively show the confusion matrix across all 16 classes. By comparing the confusion matrices of the sixteen classifiers, it is found that directions of gestures can be easily classified. In addition, most model confusions are evidently showing that the ELM-DTW method has a better gesture recognition accuracy compared with the SVM-DTW method.

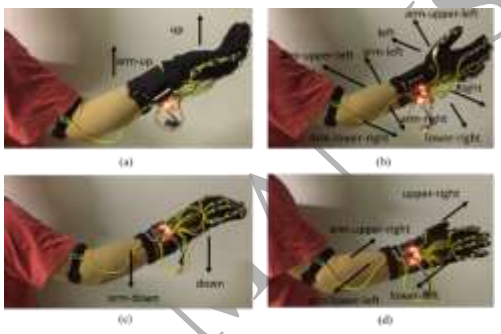

Fig. 9. The dynamic gestures of directions

Table 4. Gesture recognition accuracy and train time for different approaches

	ELM-DTW	SVM-DTW
accuracy	82.5%	78.75%

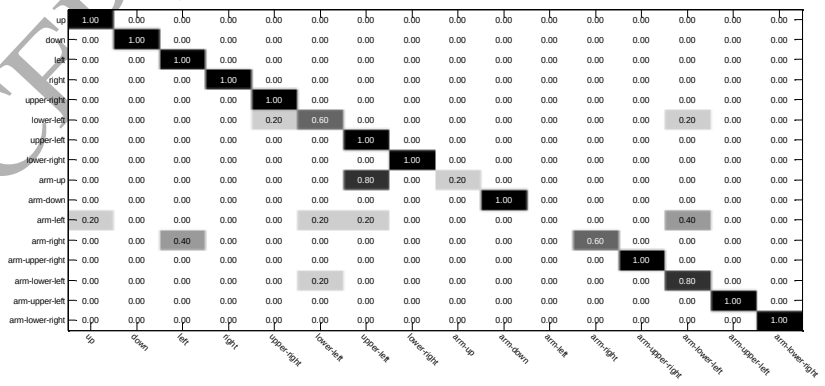

Fig. 10. The confusion matrix of ELM-DTW

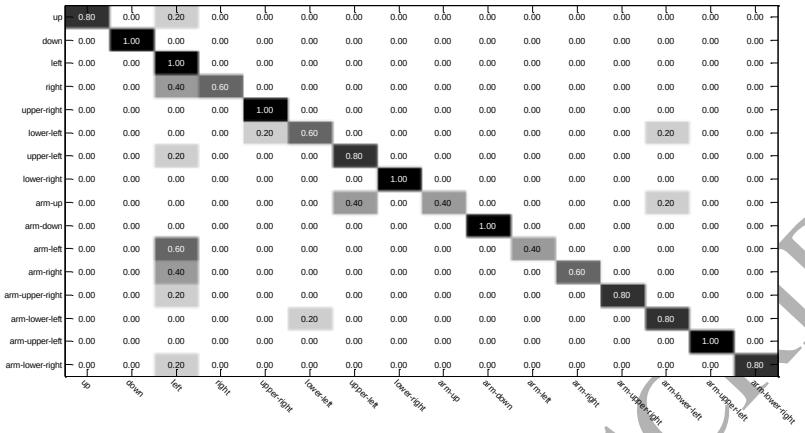

Fig. 11. The confusion matrix of SVM-DTW

5 Conclusion

We presented the design and development of a data glove for arm-hand motion capturing, and used the ELM to recognize the gestures that captured by the data glove. Commercial motion data gloves usually use high-cost motion-sensing fibers to acquire hand motion data, but in this paper we adopted the low-cost inertial and magnetic sensor to reduce cost. Meanwhile, the low-cost, low-power and light-weight IMMU is designed that superior to some commercial IMMU. Furthermore, the novel data glove is proposed based on eighteen IMMUs, which cover the whole segments of arm and hand. These advantages indicate that the proposed data glove is superior to the other data gloves. We deduced the 3D gestures capture algorithms by the multiple quaternion-based extended Kalman filters. Then the ELM-based static gestures method and dynamic gestures method are proposed. Performance evaluations verified that the proposed data glove can accurately capture the 3D gestures and the ELM-based methods can accurately recognize the gestures. In the future, it can be applied to the robotic teleoperation based on the gesture recognition.

Reference

1. S. Berman, H. Stern. Sensors for gesture recognition systems. IEEE transactions on system, man and cybernetics 2012, 42(3), 277-290.
2. D. Regazzoni, G. Vecchi, C. Rizzi, RGB cams vs RGB-D sensors: Low cost motion capture technologies performances and limitations, Journal of Manufacturing Systems, vol.33, no.4, 2014, pp.719-728.
3. L. Dipietro, A. M. Sabatini, Paolo Dario, A survey of glove-based systems and their application, IEEE Trans. Systems, Man, and Cybernetics, Part C 2008, 38(4), 461-482.
4. B. Takacs. How and why affordable virtual reality shapes the future of education. The International Journal of Virtual Reality, 2008, 7(1):53-66.

5. G. Saggio. A novel array of flex sensors for a goniometric glove. Sensors and Actuators A, 2014, 205:119-125.
6. J. M. Lambrecht, R. F. Kirsch, Miniature low-power inertial sensors: promising technology for implantable motion capture systems[J], IEEE Trans. Neural systems and rehabilitation engineering, 2014, 22(6):1138-1147.
7. Wusheng Chou, Bin Fang, Li Ding. Two-step optimal filter Design for the low-cost attitude and heading reference systems, IET Science, Measurement & Technology, 2013, 7(4):240-248.
8. Roetenberg D, Luinge HJ, Baten CTM, Veltink PH: Compensation of magnetic disturbances improves inertial and magnetic sensing of human body segment orientation. Neural Syst Rehabil Eng IEEE Trans 2005, 13(3):395–405.
9. J. M. Lambrecht, R. F. Kirsch, Miniature low-power inertial sensors: promising technology for implantable motion capture systems, IEEE Trans. Neural systems and rehabilitation engineering, 2014, 22(6):1138-1147.
10. R. Xu, S. L. Zhou, W. J. Li, MEMS accelerometer based nonspecific-user hand gesture recognition[J], IEEE sensors journal, 2012, 12(5):1166-1173.
11. Bor-Shing Lin, I-Jung Lee; Pei-Chi Hsiao; Shu-Yu Yang; Chou. Data Glove Embedded with 6-DOF Inertial Sensors for Hand Rehabilitation，2014 Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), Kitakyushu, 2014, 10:25-28.
12. Filippo Cavallo, Dario Esposito, Erika Rovini, Michela Aquilano, Maria Chiara Carrozza, Paolo Dario, Carlo Maremmani, Paolo Bongioanni. Preliminary evaluation of SensHand V1 in assessing motor skills performance in Parkinson disease, IEEE International Conference on Rehabilitation Robotics (ICORR), Seattle, USA, 2013:1-6.
13. Kortier H G, Sluiter V I, Roetenberg D, et al. Assessment of hand kinematics using inertial and magnetic sensors. Journal of Neuroengineering & Rehabilitation, 2014, 11(1):694-694.
14. Xu Y, Wang Y, Su Y, et al. Research on the Calibration Method of Micro Inertial Measurement Unit for Engineering Application. Journal of Sensors, 2016, 2016(1):1-11.
15. G. Belgioioso, A. Cenedese, G. I. Cirillo, and F. Fraccaroli. A machine learning based approach for gesture recognition from inertial measurements, IEEE Conference on Decision and Control, 2014:4899-4904.
16. O. Luzanin, and M. Plancak. Hand gesture recognition using low-budget data glove and cluster-trained probabilistic neural network, Assembly Automation, 2014, 34(34): 94-105.
17. L. Jiang, and F. Han. A Hand Gesture Recognition Method Based on SVM, Computer Aided Drafting Design & Manufacturing, 2010, 20(2):85-91.
18. G. B. Huang, D. H. Wang, Y. Lan. Extreme learning machines: a survey. International Journal of Machine Learning & Cybernetics, vol. 2, no. 2, pp. 107-122, 2011.
19. G. B. Huang, X. Ding, and H. Zhou. Optimization method based extreme learning machine for classification, *Neurocomputing*, vol. 74, no. 1-3, pp. 155-163, 2010.
20. Zhang L, Zhang D. Evolutionary Cost-Sensitive Extreme Learning Machine.[J]. IEEE Transactions on Neural Networks & Learning Systems, 2015, PP(99).
21. G. B. Huang, H. Zhou, X. Ding, and R. Zhang. Extreme learning machine for regression and multiclass classification, IEEE Transactions on Systems Man & Cybernetics Part B: Cybernetics, vol. 42, no. 2, pp. 513-29, 2012.
22. Zhang L, Zhang D. Robust Visual Knowledge Transfer via Extreme Learning Machine Based Domain Adaptation, vol. 25, no. 10, pp. 4959-4973, 2016.
23. Xsens Motion Technologies, MTi, 2015-07-1. URL, http://www.xsens.com/en/general/mti.
24. Shimmer, Shimmer website, 2015-07-1. URL, http://www.shimmersensing. com/.

25. Invense, MPU-9250 Nine-Axis MEMS MotionTracking Device, 2015-07-15. URL http://www.invensense.com/products/motion-tracking/9-axis/mpu-9250/.
26. M. Chen, S. Gonzalez, A. Vasilakos, H. Cao, and V. C. Leung, Body area networks: A survey, Mobile Networks and Applications, vol. 16, no. 2, pp. 171-193, 2011.
27. MIT Media Lab, "MIThril hardware platform," 2015-07-31. URL http://www.media.mit.edu/wearables/mithril/hardware/index.html.
28. Zhang L, Zhang D. Domain Adaptation Extreme Learning Machines for Drift Compensation in E-Nose Systems [J]. IEEE Transactions on Instrumentation & Measurement, 2015, 64(7):1790-1801.
29. P. L. Bartlett. The Sample Complexity of Pattern Classification with Neural Networks: The Size of the Weights is More Important than the Size of the Network. *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 525-536, 1998.

Bin Fang is the research assistant in the Tsinghua National Laboratory for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University in Beijing. He is interested in robotics sensor fusion and human-robot interaction.

Fuchun Sun was born in Jiangsu Province, China, in 1964. He received Ph.D degree from the Department of Computer Science and Technology, Tsinghua University, Beijing, China, in 1998. From 1998 to 2000 he was a Postdoctoral Fellow of the Department of Automation at Tsinghua University, Beijing, China. Now he is a professor in the Department of Computer Science and Technology, Tsinghua University, Beijing, China. His research interests include intelligent control, networked control system and management, neural networks, fuzzy systems,nonlinear systems and robotics.

Huaping Liu (M'07) received the Ph.D. degree from the Department of Computer Science and Technology, Tsinghua University, Beijing, China, in 2004. He is currently an Associate Professor in the Department of Computer Science and Technology at Tsinghua University. His research interests include intelligent control and robotics.

Chunfang Liu is the research assistant in the Tsinghua National Laboratory for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University in Beijing. She is interested in robotics.