



Segmentation and recognition of human motion sequences using wearable inertial sensors

Ming Guo¹ · Zhelong Wang¹

Received: 14 February 2017 / Revised: 30 October 2017 / Accepted: 21 December 2017/

Published online: 5 January 2018

© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract The application of human motion monitoring technology based on wearable inertial sensors has achieved great success in the last ten years. But now the research is mainly focused on isolated motion recognition, and there is scarce research on recognition of human motion sequences. In this paper a novel monitoring framework of human motion sequences is proposed based on wearable inertial sensors. The monitoring framework is composed of data acquisition, segmentation, and recognition stages; the main work of this paper is the last two parts. At the segmentation stage, SVD is used to perform pre-segmentation of motion sequence and its purpose is to reduce time in the segmentation process as much as possible. Then a novel similarity measure named MSH_{sim} is proposed to accomplish the fine segmentation. At the recognition stage an HMM is used to recognize the motion sequence. We use four inertial sensors to collect the human motion data. Experiments are implemented to evaluate the performance of the proposed monitoring framework, and from the experiment results, it can be seen that the proposed method may achieve better performance compared to other methods.

Keywords Wearable inertial sensors · Human motion sequence · Pre-segmentation · Fine segmentation · Motion recognition

✉ Ming Guo
guoming0537@126.com

Zhelong Wang
wangzl@dlut.edu.cn

¹ School of Control Science and Engineering, Dalian University of Technology, Dalian 116024, China

1 Introduction

Monitoring and recognition of human motion based on wearable inertial sensors is an increasingly important research field in pattern recognition and machine learning [6, 9, 31]. With the aid of wearable inertial devices like accelerometers and gyroscopes, users can obtain more information to assist themselves. Compared to video sensor-based human motion recognition, wearable inertial sensors have a number of advantages in data acquisition [13], such as easy carrying, low cost, unrestricted collection range and strong privacy preservation. Accordingly, this technique can be applied to many fields such as health care and assisted living [35, 38], prevention of diseases [32], entertainment, and promoting exercise [18, 19].

At present, the study on human motion recognition using wearable inertial sensors is mainly focused on isolated motion recognition. Although a variety of human motions have been researched [22, 26, 41], these motions have been achieved by manual segmentation which wastes a lot of time and energy. Even more important is that these motions are isolated and their correlations with each other are removed. Although the motion data and the recognition results are performed well, it is hard to apply to real life in this way because of the manual need for segmentation and the artificial nature of the data collected. Real motion data is presented in the form of a fluid sequence in practical application and human motions have randomness and disorder, which can even make manual motion identification a difficult task. Particularly, motion data is usually mixed with some useless data, so it is important to retrieve only the motion data we need from time-series data [11]. For the reasons above, a set of effective segmentation and recognition strategies for human motion sequence needs to be constructed.

Currently, research on human motion mainly uses video sensors, and the motion sequences can be divided frame-by-frame, then we may study the image associated with each frame using a common recognition algorithm [8, 10, 25, 28, 34]. In addition, some more prominent methods are proposed using video sensors. Wang et al. [39] introduced a method of motion sequence recognition based on characteristic descriptors. The authors made dynamic silhouette sequences translate into multivariate time series in the low dimension space by using tensor subspace analysis (TSA), then a support vector machine (SVM) classifier was used to classify the human motions after extracting motion descriptors in multivariate time series. Zhao et al. [43] studied gesture recognition from motion data streams based on a dynamic matching approach. The authors mainly solved two problems in this paper: one was how to monitor motions from a continuous data stream, and the other was how to recognize a gesture that had different styles. Ofli et al. [29] proposed a new definition of human motions which was described as sequences of the most informative joints (SMIJ), they just used the temporal ordering of joints to analyze motion sequences and also constructed some features to develop their method. Li et al. [24] proposed a new framework to predict long-term complex activities by mining human activity sequence patterns, the relationships between activities were established using a probabilistic suffix tree (PST), and sequential pattern mining (SPM) was used to structure the interactive information model.

Compared to recognition of human motion sequence using video sensors, only using inertial sensors to study human motion [17]. Studies based on video sensors may easily extract related motion features (such as joint angle, distance, direction and human silhouette), but it is difficult for a study using inertial sensors. Additionally, only some statistical features of the signals can be extracted, such as mean, variance, or median. Accurate trajec-

tories may be obtained using video sensors, which can help us build the motion model, but it is too difficult if only using inertial sensors to generate the motion trajectory. At present, there has less research on human motion sequences using inertial sensors. Much of the current work [4, 7, 12, 16] was focused on how to improve the performance of classification algorithms. Although they also studied motion sequences, and these papers did not consider whether there was invalid data in the motion sequences. In [2, 17], the authors only studied single-motion sequences without considering multiple-motions sequences, and the sophisticated sensors used in their work are difficult to apply in real life. The literatures on activity spotting [1, 30] systematically studied human motion sequences by using wearable inertial sensors; the first step was to extract the useful motion data in data stream, then the extracted data was recognized by using traditional classification algorithms. The main inadequacies of the papers were that the accuracy rates at the segmentation stages were relatively low.

In this paper a novel monitoring framework of human motion using wearable inertial sensors is proposed. The monitoring framework may be used to handle the human motion sequences automatically, and retrieve motion data to be recognized from motion sequences. The monitoring framework does not require manual processing, and it may save human power and material resources and time. The main contributions of our work are as follows:

- (1) We propose a systematic and automatic monitoring framework based on wearable inertial sensors. The framework mainly includes three stages: data acquisition, segmentation and recognition. Segmentation is mainly composed of pre-segmentation and fine segmentation. A hidden Markov model (HMM) is used to recognize the segmented human motion data at the recognition stage.
- (2) During pre-segmentation, singular value decomposition (SVD) is used to remove useless data as much as possible for reducing the time for the whole segmentation process.
- (3) A novel similarity measure function called multi-sensors similarity measure function (MSH_{sim}) is proposed to achieve accurate segmentation during fine segmentation.
- (4) The monitoring framework can recognize multiple human motions, and because of the use of the sliding window at the segmentation stage, online recognition can be achieved using the proposed monitoring framework.

The rest of the paper is organized as follows: in Section 2, the monitoring framework is given and a detailed introduction is given of the segmentation and recognition stages. In Section 3 the experiment and evaluation are presented; data acquisition is introduced in detail, and performance measurement and evaluation of the proposed method are discussed. This is all followed by the conclusion in the last section.

2 Methodology

The main components of the monitoring framework of human motion are shown in Fig. 1. It includes three components: data acquisition, segmentation and recognition. Data acquisition will be introduced in Section 3.1. Segmentation consists of pre-segmentation and fine-segmentation. At the recognition stage an HMM is used. In addition, for a motion sequence, the motion data to be recognized is described as *labeled data*, and the rest is called *junk data* in this paper.

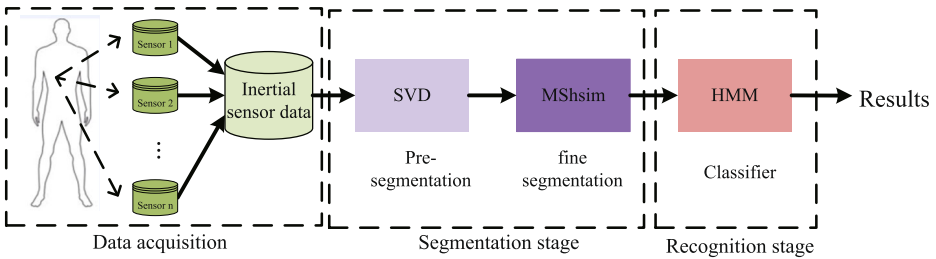


Fig. 1 Detailed structure of the human activity sequence recognition framework

2.1 Segmentation stage

2.1.1 Pre-segmentation

For segmentation of the human motion sequence, the first step is pre-segmentation as shown in Fig. 1, and its purpose is to remove the junk data. For this purpose, we use SVD combined with a sliding window. Because of this sliding window, our proposed framework has the ability of on-line processing as well.

For a real matrix $A \in R^{m \times n}$, here $m \gg n$, there must exist two orthogonal matrices $U = [u_1, u_2, \dots, u_m] \in R^m \times m$ and $V = [v_1, v_2, \dots, v_n] \in R^n \times n$, and there have,

$$A = UDV^T, \tag{1}$$

here, $D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, and $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. λ_i ($i \in \{1, 2, \dots, n\}$) represents the eigenvalues of matrix A . For (1), V is described as the right eigenvectors of real matrix A . Since V belongs to the kernel of real matrix A , and for two similar data matrices, the difference in the change of sensitivity can be obtained through calculation of the inner product of the right eigenvalues of two similar data matrices [23, 37]. Suppose v_1 and v_2 are the right eigenvectors of two similar data matrices, and α is the intersection angle of them, then we have $|v_1 * v_2| = |v_1||v_2||\cos(\alpha)|$, because v_1 and v_2 are orthogonal, so $|v_1| = |v_2| = 1$, then $|v_1 * v_2| = |\cos(\alpha)|$. When α is close to 0, $|v_1 * v_2|$ is close to 1. Particularly, when α is equal to 0 (it also means that the two data matrices are same), $|v_1 * v_2| = 1$. Most importantly, this similarity is mainly measured by the first few larger eigenvalues of the data matrices [23].

Suppose there are z sensor nodes, and the motion data from the k th sensor node is defined as matrix X_k , $k \in (1, 2, \dots, z)$. $V_k = (v_1^k, v_2^k, \dots, v_n^k)$ is the right eigenvector set of matrix X_k after SVD. In addition, the candidate matrix is Y , suppose eigenvalues of matrix Y is $(\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_n)$, and $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_n$. $\hat{V} = (\hat{v}_1, \hat{v}_2, \dots, \hat{v}_n)$ is the right eigenvector set of matrix Y after SVD. Then the segmentation function δ_k for the k th sensor node may be defined by the following equation,

$$\delta_k = \frac{1}{\omega} \sum_{i=1}^{\omega} |v_i^k * \hat{v}_i|, \tag{2}$$

and ω may be obtained by the following equation,

$$J(\omega) = \min_{\omega} \left\{ \left(\sum_{i=1}^{\omega} \hat{\lambda}_i \right) / \left(\sum_{i=1}^n \hat{\lambda}_i \right) \geq \sigma \right\}. \tag{3}$$

Here, σ is a threshold parameter. Suppose that the obtained data matrix in the k th sensor node is \tilde{X}_k after using (2), and its serial number during the sampling time is defined as $L_{pre}^k(\tilde{X}_k)$, then the final result L_{pre} for all sensor nodes can be defined as follows,

$$\begin{aligned}
 L_{pre} &= L_{pre}^1(\tilde{X}_1) \cap L_{pre}^2(\tilde{X}_2) \cap \dots \cap L_{pre}^z(\tilde{X}_z) \\
 &= \bigcap_{i=1}^z L_{pre}^i(\tilde{X}_i).
 \end{aligned}
 \tag{4}$$

Algorithm 1 Pre-segmentation algorithm

Input:

Motion sequences for z sensor nodes: $X = [X_1, X_2, \dots, X_z]$;

Candidate pattern matrix: Y ;

Incremental segment: Δ ;

Specific window: L_{win} ;

Threshold of segmentation function: T_1 ;

Output:

Motion sequences after pre-segmentation \tilde{X} ;

```

1: for  $k$ : 1 to  $z$  do
2:    $i \leftarrow 1$ ;
3:   while  $i \leq$  the size of  $X_k$  do
4:      $j \leftarrow i$ ;
5:     while  $i \leq L_{win}$  do
6:        $X_{data}$ : Data window from  $i$ th to  $(j + \Delta)$ th row;
7:       Obtain  $\delta_k$  using (2) for  $X_{data}$  and  $Y$ ;
8:       if  $\delta_k \geq T_1$  then
9:         Put  $X_{data}$  in  $\tilde{X}_k$ ;
10:      else
11:         $j \leftarrow j + 1$ ;
12:      end if
13:    end while
14:     $i \leftarrow i + 1$ ;
15:  end while
16:   $\tilde{X}_k \leftarrow unique(\tilde{X}_k)$  (Remove duplicate data);
17: end for
18: Obtain  $\tilde{X}$  using (4);
19: Return  $\tilde{X}$ ;

```

The detailed procedure for pre-segmentation is in **Algorithm 1**, T_1 is empirical parameter and it is equal to 0.85 in this paper. For a data matrix $A \in R^{m \times n}$, ($m \gg n$), the time complexity of SVD of matrix A is $O(mn^2)$, and SVD for matrix $A^T A$ can take $O(n^3)$. The right eigenvectors of A are the eigenvectors of $A^T A$ on the basis of matrix theory. So in this paper we use matrix $A^T A$ instead of matrix A to make SVD in order to save computation cost. In general, the sensitivity of this proposed method based on SVD is high according to matrix theory, and the corresponding segmentation function value (2) may be higher only when the two motions are very similar. In practice, everyone has his or her own style to perform a motion, which makes it impossible to have two fully consistent motions for different people. Even for the same person, it is impossible for the person to perform two

perfectly consistent motions due to the influence of environmental factors. In addition, if the accuracy of the inertial sensors used in the paper is not very high, it is more difficult to find a correspondence between two very similar trajectories. Two examples are given in Fig. 2. The first example includes Fig. 2a and b, and it represents that the same person performing the same motion two times. Figure 2a shows the sensor signals including acceleration signal and angular velocity signal. From Fig. 2a it can be seen that there is a difference between the two sensor signals, although the subject and the motion are the same. Figure 2b shows the corresponding eigenvalues, and the deviation of the second eigenvalues is larger, one is 313.37, the other is 580.49. For the first example, the segmentation function value is only 0.94 according to (2). The second example, including Fig. 2c and d, is obtained by different two subjects performing the same motion. From Fig. 2c it can be seen that the two curves differ. Figure 2d shows the corresponding eigenvalues, the eigenvalues of the first subject are 1367.1, 646.15, 290.95, 12.19, 6.42 and 2.81, and the eigenvalues of the second subject are 1134.5, 746.20, 215.49, 8.69, 4.62 and 1.86. For the second example, the segmentation function value is just 0.89, which has a gap compared with the result of 0.995 in [23]. Although it is difficult to accomplish the fine segmentation of the human motion sequence by this method, because of the small computational complexity mentioned above the method is chosen to complete the pre-segmentation. The objective is to reduce the time required for the whole segmentation process as much as possible.

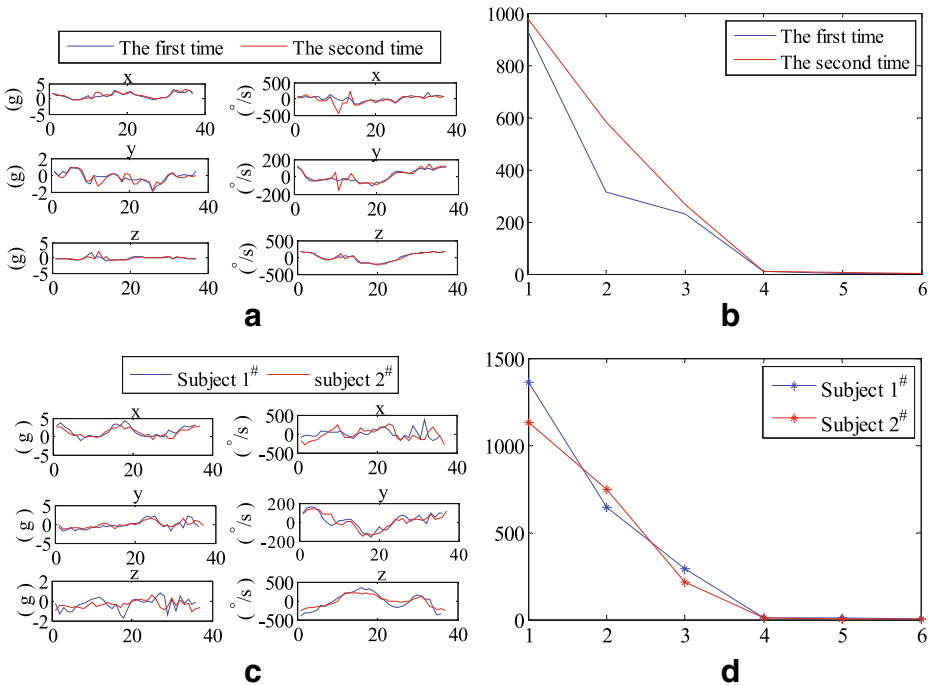


Fig. 2 Two examples for pre-segmentation. **a** The sensor signals of the activity running for the same subject. **b** The corresponding eigenvalues of the six axis sensor signals in (a). **c** The sensor signals for different two subjects. **d** The corresponding eigenvalues of the six axis sensor signals in (c)

2.1.2 Fine segmentation

After the pre-segmentation step, the new motion data can be obtained according to the (2) and (3). During fine segmentation, our purpose is to get the required motion data from the new data after pre-segmentation. We mainly use a feature similarity search between sequence data in a specific window and candidate data, and their similarity is compared using feature vectors. It is important to note that a novel similarity measure function, MSH_{sim} , is proposed to do this work. This similarity measure function is established mainly based on two aspects including the similarity measure function (H_{sim}) proposed in [42] and the characteristics of sensor data. H_{sim} is defined as the following equation,

$$H_{sim} = \frac{1}{n} \left(\sum_{i=1}^n \frac{1}{(a_i - b_i)^p} \right)^{\frac{1}{p}}, \quad (5)$$

here $A = [a_1, a_2, \dots, a_n] \in R^n$, $B = [b_1, b_2, \dots, b_n] \in R^n$ are two vectors. The function in [42] is not suitable for the present research, because it does not consider the correlation between two vectors, and this problem makes only certain features play a major role, but these features may be not very important.

Algorithm 2 Fine segmentation algorithm

Input:

Motion sequences after pre-segmentation: \tilde{X} ;
 Candidate pattern feature matrix: Y_{acc}, Y_{av} ;
 Incremental segment: Δ ;
 Specific window: L_{win} ;
 Threshold of segmentation function: T_2 ;

Output:

Motion sequences after fine-segmentation $\tilde{\tilde{X}}$;

```

1:  $i \leftarrow 1$ ;
2: while  $i \leq$  the size of  $\tilde{X}$  do
3:    $j \leftarrow i$ ;
4:   while  $i \leq L_{win}$  do
5:      $\tilde{X}_{data}$ : Data window from  $i$ th to  $(j + \Delta)$ th row;
6:     Obtain  $MSH_{sim}$  using (6) and (7);
7:     if  $\delta_k \geq T_2$  then
8:       Put  $\tilde{X}_{data}$  in  $\tilde{\tilde{X}}$ ;
9:     else
10:       $j \leftarrow j + \Delta$ ;
11:    end if
12:  end while
13:   $i \leftarrow i + \Delta$ ;
14: end while
15:  $\tilde{\tilde{X}} \leftarrow unique(\tilde{\tilde{X}})$ ;
16: Return  $\tilde{\tilde{X}}$ ;
```

Suppose there are z inertial sensor nodes, and for the inertial sensors, they can measure an acceleration signal and angular velocity signal. For candidate data, the corresponding

acceleration feature set and angular velocity feature set are $Y_{acc} = [y_1^{acc}, y_2^{acc}, \dots, y_p^{acc}]$ and $Y_{av} = [y_1^{av}, y_2^{av}, \dots, y_p^{av}]$. Acceleration feature vector is $y_i^{acc} = [y_{i1}^{acc}, y_{i2}^{acc}, \dots, y_{is}^{acc}]^T \in R^s$ and angular velocity feature vector is $y_i^{av} = [y_{i1}^{av}, y_{i2}^{av}, \dots, y_{iv}^{av}]^T \in R^v$, ($i \in (1, 2, \dots, p)$). s and v are the dimensions of acceleration features and angular velocity features, and each feature vector is composed of some basic features such as mean, variance, energy, and these features are obtained based on z sensor data fusion. For the motion sequence data obtained by pre-segmentation, suppose for one specific window whose size is greater than all candidate data length, acceleration feature vector is $x^{acc} = [x_1^{acc}, x_2^{acc}, \dots, x_s^{acc}]^T \in R^s$ and angular velocity feature vector is $x^{av} = [x_1^{av}, x_2^{av}, \dots, x_v^{av}]^T \in R^v$. Then the measure function F_{acc} for acceleration feature vector x^{acc} and measure function F_{av} for angular velocity feature vector x^{av} are as follows,

$$\begin{aligned}
 F_{acc} &= \frac{1}{ps} \sum_{i=1}^p \left(\sum_{j=1}^s \frac{1}{1+(C_i^{acc})^{-1}|x_j^{acc}-y_{ij}^{acc}|^2} \right)^{\frac{1}{2}} \\
 F_{av} &= \frac{1}{pv} \sum_{i=1}^p \left(\sum_{j=1}^v \frac{1}{1+(C_i^{av})^{-1}|x_j^{av}-y_{ij}^{av}|^2} \right)^{\frac{1}{2}}.
 \end{aligned}
 \tag{6}$$

Here, $C_i^{acc} = cov(x^{acc}, y_i^{acc})$, $C_i^{av} = cov(x^{av}, y_i^{av})$, $i \in (1, 2, \dots, p)$, and $cov(\cdot)$ represents the covariance function. So the multi-sensor similarity measure function MSH_{sim} can be defined as follows,

$$MSH_{sim} = \frac{1}{2}(F_{acc} + F_{av}).
 \tag{7}$$

According to (6), it involves the multiplication of an $n \times n$ dimensional feature matrix for z sensors, so the time complexity of MSH_{sim} is $O(z^3n^3)$. In addition, the above similarity measure function MSH_{sim} has the following characteristics, (1) MSH_{sim} may reflect the level of similarity between two data, and it can be obtained from the comparison of each dimension of the feature vectors; (2) The value range of MSH_{sim} is $[0, 1]$. If its value is higher, it shows that the two data are more similar. When MSH_{sim} is equal to 1, it indicates that these two data are equivalent. When the value of MSH_{sim} is close to 0, this indicates that the similarity of the two data is low; (3) The measure function MSH_{sim} considers the correlation between two feature vectors by using the covariance function; (4) The measure function MSH_{sim} also considers the characteristics of multi-sensors data; the acceleration data and angular velocity data are calculated separately, as they are separate scales. The detailed procedure of fine segmentation is presented in **Algorithm 2**, and in this paper T_2 is chosen in the interval $[0.96, 0.98]$.

2.2 Recognition stage

After segmenting the motion sequences, the motion data with labels needs to be recognized. In this paper, the classification algorithm we chose is an HMM. HMMs have been used in the field of human motion recognition based on wearable inertial sensors for a long time [40]. In this paper, we need to establish the HMMs for different human activities and the classification result is achieved though finding the activity sequence matching the maximum posterior probability. In addition, the features used as observations in the HMM including mean, variance, kurtosis, correlation coefficients, energy and entropy [20, 21]; the resulting number of features was 144(6 features \times 4 nodes \times 6 axes). The feature selection method we used is called robust linear discriminant analysis (RLDA) [15]. RLDA is a modified form of linear discriminant analysis (LDA), and it can solve the problem that the error exponentially

and distort discriminant analysis become larger by reestimating several smaller eigenvalues of within-class scatter matrix in LDA.

3 Experiment and evaluation

3.1 Experimental platform and data acquisition

In this paper, data acquisition is based on a multi-sensor experimental platform, which mainly includes the following components. (1) Four inertial sensor nodes, each consists of a triaxial accelerometer (ADXL325) and a triaxial gyroscope (LPR550AL), the size and the shape of the sensor node are shown in Fig. 3b. (2) One wireless receiving node, whose shape is shown in Fig. 3c. (3) A personal computer. Four sensor nodes are used to collect the motion data, and data may be uploaded to personal computer by the wireless receiving node. The data processing software is in the personal computer collecting that data.

Motion data was collected from nine subjects from our laboratory, training data was collected from five of them. The remaining four subjects were held out as a test data set. Four inertial sensor nodes were placed on the subjects, and the locations are right waist, left arm, right ankle and left thigh as shown in Fig. 3a. The sampling frequency is $50Hz$. These subjects all perform the ten same basic sports motions. These ten basic motions are walking (WA), running (RUN), turning-left waist (TLW), turning-right waist (TRW), pressing-left leg (PLL), pressing-right leg (PRL), kicking-left leg (KLL), kicking-right leg (KRL), climbing stairs (CS) and going downstairs (GD), respectively. Training data subjects are asked to

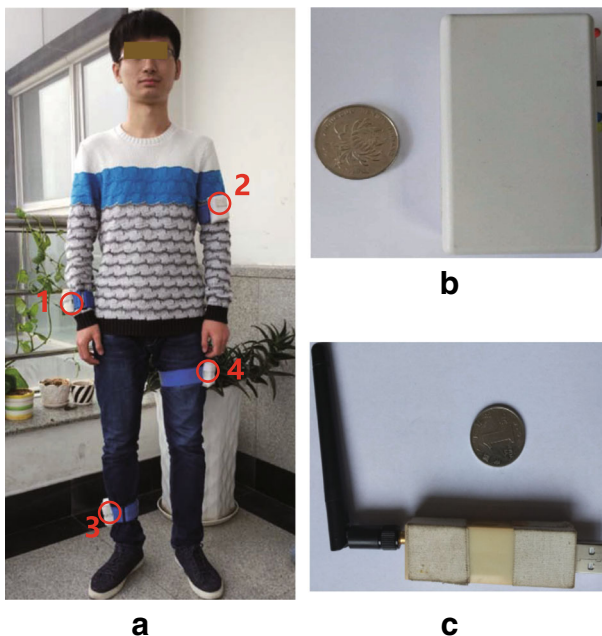


Fig. 3 The equipment in our experiment. **a** One subject with inertial sensor nodes, the positions are as follows: 1 right wrist, 2 left arm, 3 right ankle, 4 left thigh. **b** The size and shape of one inertial sensor node. **c** The shape of the receiving node

do one motion at a time, and the execution time for each motion is about 120s. Test data subjects (they are marked as *Subject 1[#]*, *Subject 2[#]*, *Subject 3[#]*, *Subject 4[#]*) are asked to carry out the corresponding motion sequences, and the execution order is shown in Fig. 4. In Fig. 4a it shows the actual execution frame of Subject 1[#], and the frame is obtained by the video, in Fig. 4b it shows the resultant acceleration curve from the right waist sensor node. Testing time is about 660s for each test data subject.

3.2 Performance measurement

In order to evaluate the performance of the proposed method, the following metrics are used [5]: *Precision*, *Recall*, *Accuracy*, and *F-score*. Suppose that TP (true positive) represents the number of positive sampled points that are classified as positive, FP (false positive) represents the number of negative sampled points that are classified as positive. FN (false negative) represents the number of positive sampled points that are classified as negative, TN (true negative) represents the number of negative sampled points that are classified as negative. The above-mentioned metrics are defined as follows,

$$Precision = \frac{TP}{TP + FP}, \tag{8}$$

$$Recall = \frac{TP}{TP + FN}, \tag{9}$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}, \tag{10}$$

$$F\text{-score} = 2 * \frac{Precision * Recall}{Precision + Recall}. \tag{11}$$

This problem is a multi-class problem, so the averages of recall and precision are used to calculate the F-score.

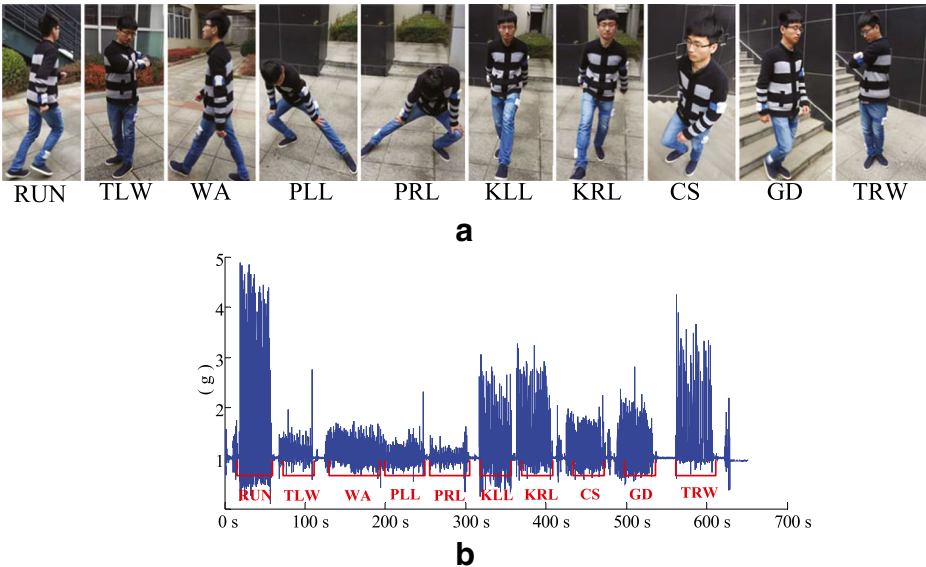


Fig. 4 The sequence of activities. **a** The actual execution frame of Subject 1[#]. **b** The resulting acceleration curve from right waist sensor node

Table 1 The comparison of running time (s) between MSH_{sim} and $SVD+MSH_{sim}$ for four subjects

	Subject 1 [#]	Subject 2 [#]	Subject 3 [#]	Subject 4 [#]
MSH_{sim}	3472	3499	3053	3777
$SVD+MSH_{sim}$	1745	1727	1604	1863

3.3 Performance evaluation for segmentation

In this section, we evaluate the running time for segmentation for multi-sensor data. It shows the comparison of running time between MSH_{sim} and $SVD+MSH_{sim}$ for four subjects in Table 1. From the table it can be seen that, running time is almost more than double if we only use the fine segmentation method MSH_{sim} without SVD. This demonstrates the effectiveness of the per-segmentation method based on SVD proposed in this paper.

In order to evaluate the segmentation performance, four subjects are required to perform the motion sequences mentioned in section III-A, and the actual results are shown in Fig. 5. Figure 5a shows the segmentation result of Subject 1[#], the retrieved motions are represented as labeled data, and the rest are defined as junk data. The blue curve in the first box of Fig. 5a represents the real situation, and it is obtained by using the video, the red curve of Fig. 5a is the result of data segmentation by using the proposed method. Similarly, the results for the remaining three subjects can be obtained. From the figure, it can be seen that the result of Subject 1[#] is consistent with expectations. But for Subject 2[#], this result is not good compared with the actual condition. Overall, this figure shows that most of the labeled

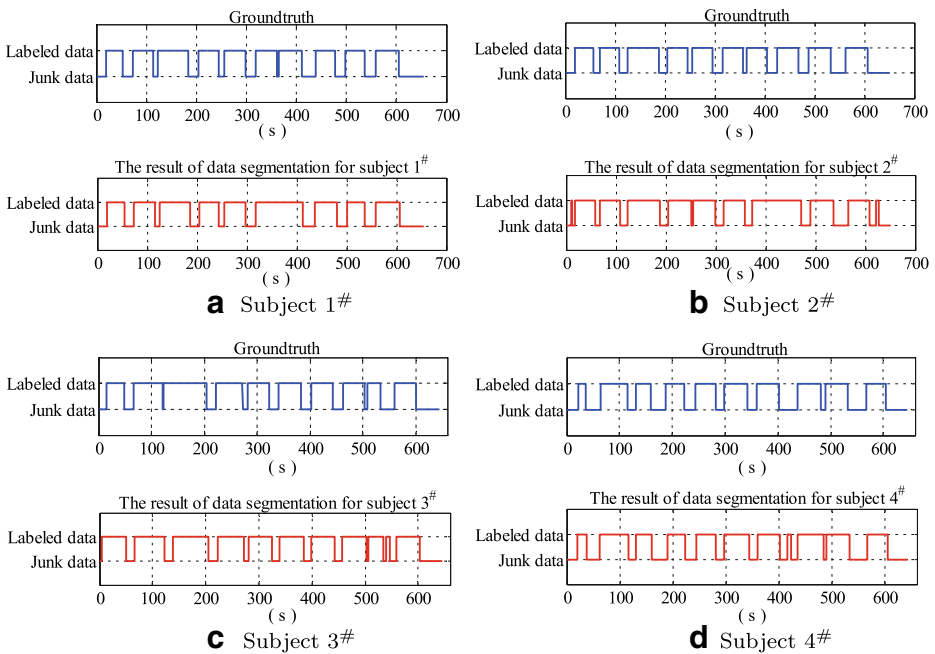


Fig. 5 The results of data segmentation for four subjects by using the proposed segmentation method. The blue curves represent the real situations, and the red curves are the segmentation results

Table 2 The confusion matrices of four subjects during segmentation. LD represents *labeled data*, JD represents *junk data*

	LD	JD	Total	Recall
(a) Subject 1 [#]				
LD	21055	455	21510	97.88%
JD	991	10202	11193	91.15%
Total	22046	10657	32703	
Precision	95.60%	95.73%		
(b) Subject 2 [#]				
LD	20864	784	21648	96.38%
JD	2752	8138	10890	74.73%
Total	23616	8922	32538	
Precision	88.35%	91.21%		
(c) Subject 3 [#]				
LD	21740	696	22436	96.90%
JD	2186	7597	9783	77.66%
Total	23926	8293	32219 *	
Precision	90.86%	91.96%		
(d) Subject 4 [#]				
LD	18635	250	18885	98.68%
JD	1396	11927	13323	89.52%
Total	20031	12177	32208	
Precision	93.03%	97.95%		

data may be retrieved by using the proposed method, although there may be some errors for the four subjects.

In order to show the segmentation effects more clearly, the confusion matrices combining with precision and recall of the four subjects are given in Table 2. For Subject 1[#], the precision and recall of labeled data are 95.60% and 97.88% respectively, and both of them are high. The precision and recall of junk data are also high at 95.73% and 91.15% respectively, which shows that the proposed segmentation method is effective. For Subject 2[#], the precision and recall of labeled data are 88.35% and 96.38%, and the precision and recall of junk data are 91.21% and 74.73%. The recall of junk data is not high, it shows the junk data in excess of 20% is wrongly assigned as labeled data. The same situation also appears for Subject 3[#], the precision and recall of labeled data are 90.86% and 96.90% respectively, and the precision and recall of junk data are 91.96% and 77.66%. For Subject 4[#], the results are very good like the Subject 1[#]. In summary, from the Table 2 it can be seen that the precisions and recalls for all of the subjects are good, again, this shows that the segmentation method proposed in the paper is effective. Table 3 gives the detailed segmentation results of four metrics by using SVD+MSH_{sim}.

To express the superiority of the proposed segmentation method based on SVD and MSH_{sim} in the paper, we make the comparisons between our method and other methods including H_{sim} and Euclidean distance [1, 3]. Euclidean distance is a very useful similarity measuring technique, and it can be defined as follows,

$$Euclidean = ((x - y)(x - y)^T)^{\frac{1}{2}},$$

Table 3 The detailed results (%) of Precision, Recall, Accuracy and F-score by using SVD+MSH_{sim} at segmentation stage, \bar{P} represents the average of precision values and \bar{R} represents the average of recall values

	\bar{P}	\bar{R}	Accuracy	F-score
Subject 1 [#]	95.67	94.52	95.58	95.09
Subject 2 [#]	89.78	85.56	89.13	87.62
Subject 3 [#]	91.41	87.28	91.05	89.30
Subject 4 [#]	95.49	94.10	94.89	94.79

where $x \in R^n$ and $y \in R^n$ are two feature vectors. Figure 6 gives the accuracy rate of segmentation between different segmentation methods including H_{sim}, Euclidean distance and our method. From Fig. 6 it can be seen that MSH_{sim} outperforms the other two algorithms for four subjects. For Subject 1[#], the accuracy rate of MSH_{sim} is 95.58%, the other two methods are 92.38% and 79.50%. For Subject 2[#], the result of our method is just 89.13%, which is not still exceeded by other methods, and their results are 86.74% and 77.42%. For Subject 3[#], the results are 91.05%, 88.93% and 78.79% corresponding to MSH_{sim}, H_{sim} and Euclidean distance. The accuracy rates are 94.89%, 84.32% and 75.98% for Subject 4[#]. As you can see from the results, both MSH_{sim} and H_{sim} can obtain better performance than that of Euclidean distance. That is because the first two methods calculate the level of similarity from each dimension of the feature vectors, but the similarity measure by using Euclidean distance is calculated holistically. This will lead to greater error, especially for the acceleration data and angular velocity data that have different orders of magnitude. In addition, our method MSH_{sim} is slightly better than H_{sim}, because we fully consider the characteristics of the sensor data, and use the covariance function to solve the correlation among axes. Because the acceleration data and angular velocity data have different dimensions, they are processed separately.

3.4 Performance evaluation for recognition

After segmentation retrieved data is classified into the ten kinds of human motions introduced in section III-A. The recognition algorithm we chose is HMM because of its ability to process sequences. Figure 7 gives the detailed classification results of four subjects. The red curves represent the types of motion for recognition, and the blue curves represent the actual acquisition data. From the four pictures in Fig. 7, the two curves in each picture are basically the same. In detail, the red curve and blue curve in Fig. 7a are closer than the other three pictures except TLW and TRW, which suggests that the classification result of Subject 1[#] is the best among these four subjects. The overlap of two curves in Fig. 7d is not well especially for KLL, and this shows that the classification effect of Subject 4[#] is poor. Next the concrete results are given in order to express the classification effects better.

Table 4 gives the corresponding confusion matrices for four subjects. For Subject 1[#] (Table 4(a)), it can be seen that the recalls and the precisions of TLW and TRW are relatively low compared with other motions, and they are not more than 90%. The reason is that these two motions are easy to confuse, which is consistent with what is shown in Fig. 7a. The classification results of other motions in Table 4(a) are relatively better. For Subject 2[#] (Table 4(b)), the precisions of four motions including PRL, KLL, KRL and GD are much worse than other motions, and their precisions are 79.61%, 79.08%, 60.26% and 81.18%, respectively. The recalls of TRW and KRL are just 79.01% and 76.15%, respectively. The reason is principally because the error is larger in the segmentation stage for Subject 2[#]. The

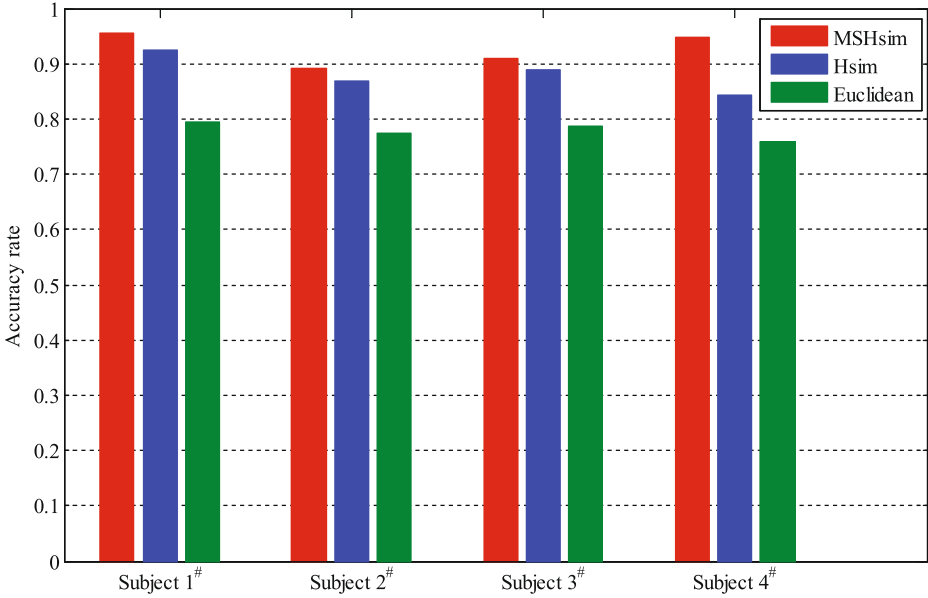


Fig. 6 The accuracy rate of segmentation between different segmentation methods including H_{sim} , Euclidean distance and MSH_{sim}

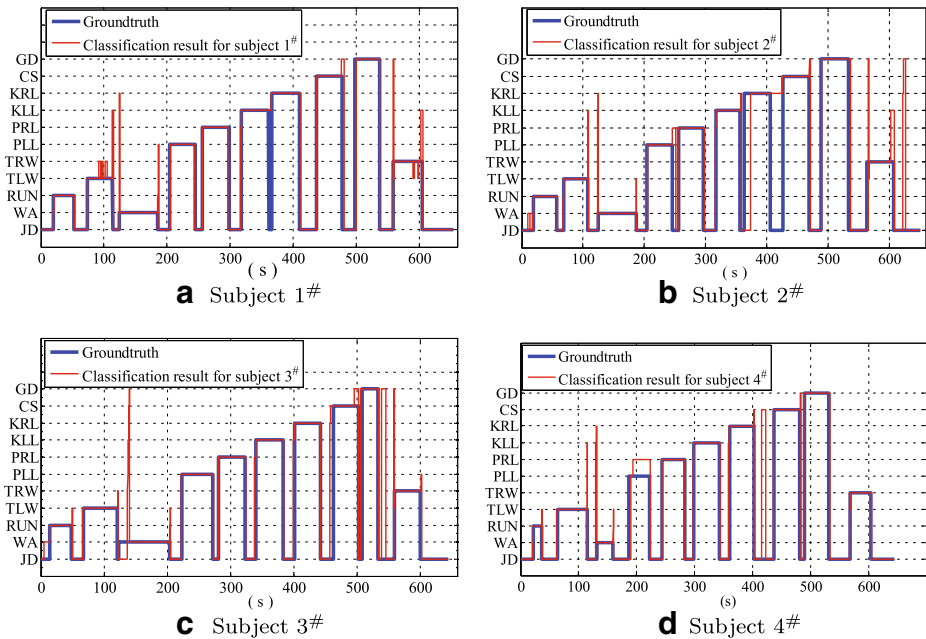


Fig. 7 The classification results of four subjects compared with the actual situations, where the red curve of each picture represents the classification result, and the blue curve represents the corresponding actual situation. JD represents *junk data*

Table 4 The confusion matrices of four subjects in recognition stage

	WA	RUN	TLW	TRW	PLL	PRL	KLL	KRL	CS	GD	JD	Total	Recall (%)
(a) Subject 1[#]													
WA	2980	0	0	0	0	0	13	25	0	0	25	3034	97.93
RUN	25	1587	0	0	0	0	0	0	0	0	7	1619	98.02
TLW	0	0	1540	400	0	0	53	0	0	0	7	1993	77.27
TRW	0	0	175	1975	0	0	118	0	0	29	0	2297	85.98
PLL	0	0	0	0	1903	0	0	0	0	0	157	2060	92.38
PRL	0	0	0	0	0	2033	0	0	0	0	110	2143	94.87
KLL	0	0	0	0	0	0	2155	0	0	0	0	2155	100
KRL	0	0	0	0	0	0	0	2233	0	0	0	2233	100
CS	0	0	0	0	0	0	0	0	1921	77	0	1998	96.15
GD	0	0	0	0	0	0	0	0	0	1813	156	1969	92.08
JD	45	113	35	0	47	42	382	142	4	181	10202	11193	91.15
Total	3050	1700	1750	2375	1950	2075	2721	2400	1925	2100	10657	32703	
Precision (%)	97.70	93.35	88.00	83.16	97.59	97.98	79.20	93.04	99.79	86.33	95.73		
(b) Subject 2[#]													
WA	3059	25	20	0	0	0	0	0	0	0	0	3104	98.55
RUN	0	1882	0	0	0	0	0	0	0	0	0	1882	100
TLW	0	0	1919	0	0	0	12	0	0	0	0	1931	99.38
TRW	0	0	25	1750	0	0	274	0	0	11	155	2215	79.01
PLL	0	0	3	0	1975	20	0	0	0	0	67	2065	95.64
PRL	0	0	0	0	0	2030	0	0	0	0	0	2030	100
KLL	0	0	0	0	0	0	1977	0	0	0	0	1977	100
KRL	0	0	0	0	0	0	0	1609	0	0	504	2113	76.15
CS	0	0	0	0	0	0	0	0	2125	0	0	2125	100
GD	0	0	25	0	0	0	0	0	0	2148	58	2206	97.37
JD	216	143	8	0	0	500	237	1061	100	487	8138	10890	74.73
Total	3275	2050	1975	1750	1975	2550	2500	2670	2225	2646	8922	32538	
Precision (%)	93.40	91.80	97.16	100	100	79.61	79.08	60.26	95.51	81.18	91.21		
(c) Subject 3[#]													
WA	3176	0	0	6	25	25	25	25	25	25	696	4028	78.85
RUN	13	1693	0	0	0	0	0	0	0	0	0	1706	99.24
TLW	0	0	2681	0	0	0	0	0	0	0	0	2681	100
TRW	0	0	21	2036	0	0	0	0	0	0	0	2057	98.98
PLL	0	0	0	0	2482	0	0	0	0	0	0	2482	100
PRL	0	0	0	0	0	2064	0	0	0	0	0	2064	100
KLL	0	0	0	0	0	0	2135	0	0	0	0	2135	100
KRL	0	0	0	0	0	0	9	2038	0	0	0	2047	99.56
CS	0	0	0	0	0	0	0	0	1633	329	0	1962	83.23
GD	0	0	0	0	0	0	0	0	0	1274	0	0	100
JD	486	82	123	109	68	186	206	62	242	622	7597	9783	77.66
Total	3675	1775	2825	2151	2575	2275	2375	2125	1900	2250	8293	32219	
Precision (%)	86.42	95.38	94.90	94.65	96.39	90.73	89.89	95.91	85.95	56.62	91.61		

Table 4 (continued)

	WA	RUN	TLW	TRW	PLL	PRL	KLL	KRL	CS	GD	JD	Total	Recall (%)
(d) Subject 4 [#]													
WA	1403	0	0	0	0	0	0	0	0	0	0	1403	100
RUN	25	707	0	0	0	0	0	0	0	0	7	707	100
TLW	0	0	2592	0	0	0	6	0	0	0	0	2598	97.77
TRW	0	0	69	1768	0	0	0	0	0	0	0	1837	96.24
PLL	0	0	2	0	150	1433	0	0	0	0	191	1776	8.45
PRL	0	0	2	0	0	1878	0	0	0	0	59	1937	96.95
KLL	0	0	0	0	0	0	2237	0	0	0	0	2237	100
KRL	0	0	0	0	0	0	0	2071	0	0	0	2017	100
CS	0	0	0	0	0	0	0	0	2214	0	0	2214	100
GD	0	0	0	0	0	0	0	0	0	2105	0	2105	100
JD	72	68	137	63	0	139	182	79	461	195	11927	13323	89.52
Total	1475	775	2800	1831	150	3450	2425	2150	2675	2300	12177	32208	
Precision (%)	95.12	91.23	92.57	96.56	100	54.43	92.25	96.33	82.77	91.52	97.95		

JD represents *junk data*

precision of GD is very poor for Subject 3[#], because some data in CS and some junk data are erroneously classified as GD, but for other motions their results are good. For Subject 4[#], most of the data in PLL are erroneously classified as PRL. As a result, the recall of PLL is very low at only 8.45%, and the precision of PRL is only 54.43%, which is consistent with that shown in Fig. 7d. In short, we can achieve good results for motion sequence classification using an HMM. Table 5 gives the detailed recognition results of four metrics by using an HMM.

Figure 8 shows the comparison of recognition results between the HMM and other four classification algorithms including K-nearest neighbor (KNN) [36], naive Bayes (NB) [33], softmax regression (SR) [14], and linear discriminant analysis classifier (LDAC) [27]. From Fig. 8 it can be seen that the HMM classification algorithm can achieve the best accuracy rates for four subjects, 92.75%, 87.75%, 89.42% and 90.21%, respectively. The results of both NB and LDAC are relatively stable, and their accuracy rates are more than 80% for four subjects but all are lower than that by using the HMM. The KNN result is relatively low for each subject, and the effect of SR is also not satisfying from Fig. 8. This demonstrates that the classification algorithm chosen in this paper is effective.

Table 5 The detailed results (%) of Precision, Recall, Accuracy, and F-score by using an HMM for the recognition stage

	\bar{P}	\bar{R}	Accuracy	F-score
Subject 1 [#]	91.99	93.26	92.75	92.62
Subject 2 [#]	88.11	92.80	87.75	90.05
Subject 3 [#]	88.95	94.32	89.42	91.56
Subject 4 [#]	90.07	89.90	90.21	89.98

\bar{P} represents the average of precision values and \bar{R} represents the average of recall values

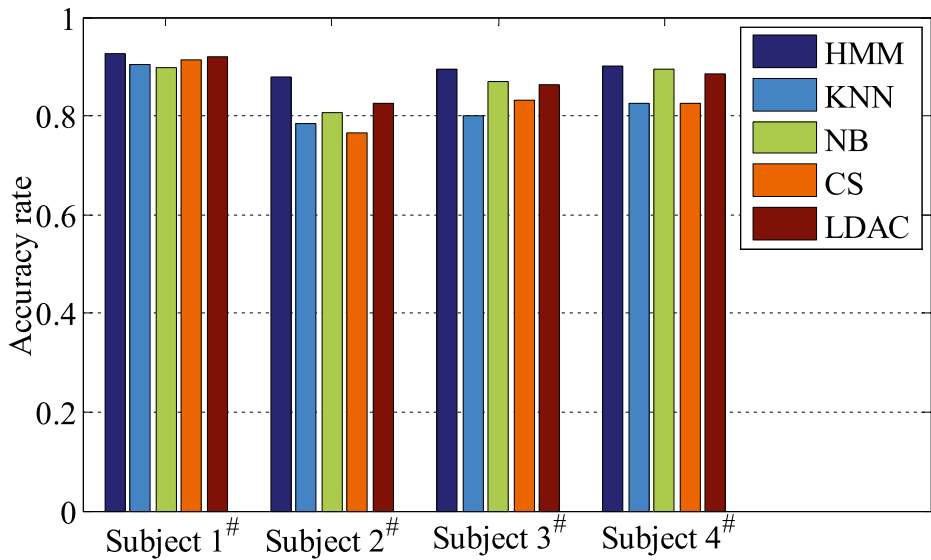


Fig. 8 The comparison of recognition accuracy rates between the HMM and four other classification algorithms including KNN, NB, SR and LDAC

4 Conclusion

In order to realize the segmentation and recognition of human motion sequences using wearable inertial sensors, a novel recognition framework is proposed; this is mainly composed of a data acquisition stage, segmentation stage, and recognition stage. Four inertial sensors are used to collect human motion data. The segmentation stage is divided into two steps including pre-segmentation and fine segmentation. At the pre-segmentation step SVD is used to delete the junk data in the human motion sequences. We also proposed a novel similarity measure, MSH_{sim} , to realize the accurate segmentation of motion sequence during fine segmentation. An HMM was used during the recognition stage. Motions sequences from four test data subjects are used to validate the proposed methods. The experimental results demonstrate the effectiveness of the proposed framework.

In our future work, there are still some problems that need further research. First, the on-line recognition should be realized if we want to use this method in real life, and the performance of inertial sensor will be improved to achieve long-term monitoring of human motions, and we also consider to establish the monitoring system based on cloud computing in future work. Second, time consumption is still great because of multi-sensors fusion, it needs to be reduced by means of the most simplified algorithm. Third, in future work some of the more complex motion sequences need to be studied.

Acknowledgements This work was supported by National Natural Science Foundation of China under Grant No.61473058, Fundamental Research Funds for the Central Universities (DUT15ZD114) and Project Funded by China Postdoctoral Science Foundation (2017M621131). The authors gratefully acknowledge the assistance of Mark V. Albert in correcting English language.

References

1. Amft O (2011) Self-taught learning for activity spotting in on-body motion sensor data. In: 2011 15th annual international symposium on wearable computers. IEEE, pp 83–86
2. Amft O, Bannach D, Pirkl G, Kreil M, Lukowicz P (2010, April) Towards wearable sensing-based assessment of fluid intake. In: PerCom Workshops, pp 298–303
3. Amft O, Troster G (2008) Recognition of dietary activity events using on-body sensors. *Artif Intell Med* 42(2):121–136
4. Andreu J, Angelou P (2010) Real-time human activity recognition from wireless sensors using evolving fuzzy systems. In: 2010 IEEE International Conference on Fuzzy Systems (FUZZ). IEEE, pp 1–8
5. Bama SS, Ahmed ML, Saravanan A (2015) A survey on performance evaluation measures for information retrieval system
6. Bhimani J, Mi N, Leeser M, Yang Z (2017) FIM: performance prediction for parallel computation in iterative data processing applications. In: 2017 IEEE 10th International Conference on Cloud Computing (CLOUD). IEEE, pp 359–366
7. Blanke U, Schiele B (2009) Daily routine recognition through activity spotting. In: International symposium on location-and context-awareness. Springer, Berlin, pp 192–206
8. Candès E, Li X, Ma Y, Wright J (2010) Robust principal component analysis?: recovering low-rank matrices from sparse errors. 8(1):201–204
9. Chen L, Hoey J, Nugent CD, Cook DJ, Yu Z (2012) Sensor-based activity recognition. *IEEE Trans Syst Man Cybern Part C Appl Rev* 42(6):790–808
10. Elhamifar E, Vidal R (2013) Sparse subspace clustering: algorithm, theory, and applications. *IEEE Trans Pattern Anal Mach Intell* 35(11):2765–2781
11. Ghaleb FF, Youness EA, Elmezain M, Dewdar FS (2015) Vision-based hand gesture spotting and recognition using CRF and SVM. *J Softw Eng Appl* 8(07):313
12. Ghassemzadeh H, Guenterberg E, Ostadabbas S, Jafari R (2009) A motion sequence fusion technique based on pca for activity analysis in body sensor networks. In: 2009 annual international conference of the IEEE engineering in medicine and biology society. IEEE, pp 3146–3149
13. Guan D, Ma T, Yuan W, Lee YK, Jehad Sarkar AM (2011) Review of sensor-based activity recognition systems. *IETE Tech Rev* 28(5):418–433
14. Guo J, Xie X, Bie R, Sun L (2014) Structural health monitoring by using a sparse coding-based deep learning algorithm with wireless sensor networks. *Pers Ubiquit Comput* 18(8):1977–1987
15. Guo M, Wang Z (2015) A feature extraction method for human action recognition using body-worn inertial sensors. In: 2015 IEEE 19th International Conference on Computer Supported Cooperative Work in Design (CSCWD). IEEE, pp 576–581
16. Hammerla NY, Halloran S, Ploetz T (2016) Deep, Convolutional, and Recurrent Models for Human Activity Recognition using Wearables. arXiv:1604.08880
17. Junker H, Amft O, Lukowicz P, Trster G (2008) Gesture spotting with body-worn inertial sensors to detect user activities. *Pattern Recogn* 41(6):2010–2024
18. Kunze K, Barry M, Heinz EA, Lukowicz P, Majoe D, Gutknecht J (2006) Towards recognizing tai chi-an initial experiment using wearable sensors. In: 2006 3rd International Forum on Applied Wearable Computing (IFAWC). VDE, pp 1–6
19. Ladha C, Hammerla NY, Olivier P, Plotz T (2013) ClimbAX: skill assessment for climbing enthusiasts. In: Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing. ACM, pp 235–244
20. Lara OD, Labrador MA (2013) A survey on human activity recognition using wearable sensors. *IEEE Commun Surv Tutor* 15(3):1192–1209
21. Lee MW, Khan AM, Kim TS (2011) A single tri-axial accelerometer-based real-time personal life log system capable of human activity recognition and exercise information generation. *Pers Ubiquit Comput* 15(8):887–898
22. Leutheuser H, Schuldhuis D, Eskofier BM (2013) Hierarchical, multi-sensor based classification of daily life activities: comparison with state-of-the-art algorithms using a benchmark dataset. *PloS One* 8(10):e75196
23. Li C, Zheng SQ, Prabhakaran B (2007) Segmentation and recognition of motion streams by similarity search. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 3(3):16
24. Li K, Fu Y (2014) Prediction of human activity by discovering temporal sequence patterns. *IEEE Trans Pattern Anal Mach Intell* 36(8):1644–1657
25. Liu G, Lin Z, Yan S, Sun J, Yu Y, Ma Y (2013) Robust recovery of subspace structures by low-rank representation. *IEEE Trans Pattern Anal Mach Intell* 35(1):171–184

26. Lu Y, Wei Y, Liu L (2016) Towards unsupervised physical activity recognition using smartphone accelerometers[J]. *Multimed Tool Appl* 2016:1–19
27. Murtaza M, Sharif M, Raza M, Shah J (2014) Face recognition using adaptive margin fishers criterion and linear discriminant analysis. *Int Arab J Inform Technol* 11(2):1–11
28. Ni B, Wang G, Moulin P (2013) Rgb-d-hudaact: a color-depth video database for human daily activity recognition. In: *Consumer Depth Cameras for Computer Vision*. Springer, London, pp 193–208
29. Ofli F, Chaudhry R, Kurillo G, Vidal R, Bajcsy R (2014) Sequence of the most informative joints (smij): a new representation for human skeletal action recognition. *J Vis Commun Image Represent* 25(1):24–38
30. Ogris G, Stiefmeier T, Lukowicz P, Troster G (2008) Using a complex multi-modal on-body sensor system for activity spotting. In: *2008 12th IEEE international symposium on wearable computers*. IEEE, pp 55–62
31. Ordonez FJ, de Toledo P, Sanchis A (2015) Sensor-based bayesian detection of anomalous living patterns in a home setting. *Pers Ubiquit Comput* 19(2):259–270
32. Paradiso R, Loriga G, Taccini N (2005) A wearable health care system based on knitted integrated sensors. *IEEE Trans Inf Technol Biomed* 9(3):337–344
33. Rish I (2001) An empirical study of the naive Bayes classifier. In: *IJCAI 2001 workshop on empirical methods in artificial intelligence*, vol 3, No. 22. IBM, New York, pp 41–46
34. Shi Q, Cheng L, Wang L, Smola A (2011) Human action segmentation and recognition using discriminative semi-markov models. *Int J Comput Vis* 93(1):22–32
35. Singla G, Cook DJ, Schmitter-Edgecombe M (2010) Recognizing independent and joint activities among multiple residents in smart environments. *J Ambient Intell Humaniz Comput* 1(1):57–63
36. Song Y, Huang J, Zhou D, Zha H, Giles CL (2007) Iknn: informative k-nearest neighbor pattern classification. In: *European conference on principles of data mining and knowledge discovery*. Springer, Berlin, pp 248–264
37. Stewart GW (1973) Error and perturbation bounds for subspaces associated with certain eigenvalue problems. *SIAM Rev* 15(4):727–764
38. Van Kasteren TLM, Englebienne G, Krose BJ (2010) An activity monitoring system for elderly care using generative and discriminative models. *Pers Ubiquit Comput* 14(6):489–498
39. Wang L, Wang X, Leckie C, Ramamohanarao K (2008) Characteristic-based descriptors for motion sequence recognition. In: *Pacific-Asia conference on knowledge discovery and data mining*. Springer, Berlin, pp 369–380
40. Wang Z, Guo M, Zhao C (2016) Badminton stroke recognition based on body sensor networks. *IEEE Trans Human-Machine Syst* 46(5):769–775
41. Wu D, Wang Z, Chen Y, Zhao H (2016) Mixed-kernel based weighted extreme learning machine for inertial sensor based human activity recognition with imbalanced dataset. *Neurocomputing* 190:35–49
42. Yang F, Zhu Y, Shi BL (2004) An efficient method for similarity search on quantitative transaction data. *J Comput Res Develop* 41(2):361–368
43. Zhao X, Li X, Pang C, Zhu X, Sheng QZ (2013) Online human gesture recognition from motion data streams. In: *Proceedings of the 21st ACM international conference on multimedia*. ACM, pp 23–32



Ming Guo received the B.S. and M.S. degrees in Linyi University and Northeastern University, China, in 2010 and 2012, respectively. He is currently pursuing the Ph.D. degree in control theory and control engineering with the School of Control Science and Engineering, Dalian University of Technology, Dalian, China. His research interests include pattern recognition, body sensor networks and bioinformatics.



Zhelong Wang received the B.S. and M.S. degrees in automatic control from the Dalian University of Technology, Dalian, China, in 1996 and 1999, respectively, and the Ph.D. degree in robotics from the University of Durham, U.K., in 2003. In 2004, he joined the School of Electronic and Information Engineering, Dalian University of Technology, where he is currently a Professor and a Ph.D. Supervisor. His research interests include robotics, intelligent control, mechatronics, body sensor networks, and machine learning. He received a Research Award from the Audi Foundation in 2008, the Dalian Excellent IT Teacher Award in 2008, and the first prize of the Liaoning Province Natural Science Award (Academic Paper Program) in 2009. He was elected into Liaoning Province Hundred, Thousand, and Ten Thousand Talent Project in 2009.