Human Motion Recognition Based On Inertial Sensor

Mengmeng Xing, Guohui Wei, Hui Cao*, Feng Yang, Yongqi Nie and Jing Liu

College of Science and Technology Shandong University of Traditional Chinese Medicine

*Corresponding author e-mail: caohui63@163.com, 13256688200@163.com, bmie530@163.com, ayangfeng163@163.com, 13371566657@163.com, 13954125302@163.com

Abstract. Human activity recognition (HAR) has attracted considerable research attentions from all walks of life due to its application in human-computer interaction, smart medical treatment and smart home health care. There are various ways of using different motion capture methods for HAR. Among which, HAR based on inertial sensor signals has recently emerged as a challenging and hot research topic. In this paper, we presented a inertial sensor-based method for HAR. Firstly, the variance, mean, information entropy, peak, etc. were extracted from the inertial data as robust features of the actions. Then, the robust features were further processed by Principal Component Analysis (PCA) to reduce the dimensions of features. Finally, the features with different actions label were used to train the Extreme Learning Machine (ELM) and Support Vector Machine (SVM) model. The result can achieve an mean accuracy of 83.49%, it also verified the effectiveness of our method.

1. Introduction

Human body movements can reflect abundant information, such as the purpose of behavior, health status, etc. In recent years, human motion information has become an elegant research field for its hidden value. HAR was first developed based on computer vision. The effective recognition of human action from computer can be realized: Smart Medical [1], Sports Training [2]. Until now, Scholars have established multiple motion databases through camera recording, including KTH database, Weizmann database, INRIA XMAS database, etc.

Though computer vision recognition system has been very popular for HAR. However, it is difficult to achieve the accuracy of human visual perception. Despite a lot of efforts have made by researchers, HAR based on computer vision still faces many limitations, such as: (1) the limitation of motion range, (2) the disturbance of external environment, and (3) the dual complexity of human motion in time and space. On the contrary, inertial sensors can overcome the above limitation. Now with the rapid development of Micro Electro Mechanical Systems (MEMS), inertial sensor can be attached to human joints as wearable equipment. The micromechanical motion capture using wearable sensors has gradually attracted the attention of researchers. Mumtaz et al. [3] developed a pedestrian navigation application by using a low-cost wireless inertial measurement unit and compared it with the sensor kit in the Samsung galaxy s5 phone. The experiment found that it can accurately track the trend of angle and acceleration. The inertial sensor is capable of capturing motion due to its high sensitivity,
it can also be used for the measurement of joint angles. Now Seel et al. [4] proposed a new joint axis position recognition method, using only gyroscopes and accelerometers. It can achieve the measurement of flexion and extension joint. Wittmann et al. [5] established a virtual reality system based on inertial measurement unit. Chronic stroke patients can complete upper limb training without supervision in this system. The experiment results showed that this unsupervised family treatment plan was feasible. Robert-Lachaine et al. [6] verified the validity of inertial sensor in joints analysis and evaluated the impact of task complexity and persistence on the inertial measurement unit. Inertial sensors are now widely used in the process of capturing motion trajectories.

In last few decades, the HAR system has mainly contained three main parts: motion capture, feature extraction and motion recognition. Recently, feature extraction have attracted many activity recognition researchers as it can reduce computational complexity and increase accuracy. Tao et al. [7] proposed the concept of motion pose elements, which can represent human body level from limbs to whole body, and developed a middle-level feature learning framework based on motion pose elements to extract skeleton data features for motion recognition. Ofli et al. [8] proposed joint sequence information (SMIJ) to select features such as bone joint angle and angular velocity at a specific time for action representation. For actions characterized by inertial sensors (e.g., accelerometers and gyroscopes), Guerra et al. [9] used statistical tools based on sliding windows to represent motion data, such as mean, variance, maximum, minimum, and root mean square. In the pattern recognition, SVM and ELM are the most classic classifiers. For example, Lin et al. [10] combined SVM with Deep Convolutional Neural Networks (DCNN), which could automatically recognize multi-channel microseismic waveforms with a recognition rate of 98.18%. Anguita et al. [11] presented a system for human physical Activity Recognition (AR). The system used SVM to classify 6 actions, and the classification accuracy rate reached 89%. In our experiment, SVM and ELM were also used as classifiers to classify human movements.

2. EXPERIMENTS DETAIL

Fig.1 shows the basic flowchart of our proposed method, which mainly consists of three steps: motion capture, feature extraction, and motion recognition. For the motion capture step, the inertial sensor data were collected as input of the HAR system. The second step starts with removing noise by using smoothing filter. After removing noise, it did statistical analysis on fixed-size sliding windows over the time-sequential signals to obtain robust features. The third step distinguished activities from the robust features by ELM and SVM.

![Flowchart of the proposed HAR method.](image-url)
2.1. Inertial signal processing

The human motion data in this experiment was obtained from UTD-MHAD dataset. It is a multi-modal human behavior recognition data set published by Chen et al. at the University of Texas [12]. The dataset contains 27 kinds of action data which are composed with 861 data sequences, the specific categories of actions are shown in Tab. 1.

The sampling rate of the inertial sensor signals is 50 Hz, accelerometer and gyroscope sensors synchronously record datas on the X, Y, and Z axes. Sensor jitter inevitably occurred during human motion signal acquisition, we used a smoothing filter to reduce the noise of the original signal, 0.1 seconds is considered to be the optimal window width for smoothing filter. The raw data and the data after noise reduction are both shown in Fig. 2.

Table 1. 27 actions in the UTD-MHAD data set

| Inertial sensor on right wrist | Action type | Action |
|--------------------------------|-------------|--------|
| **Sports movement**           |             | basketball shoot bowling boxing baseball swing from right tennis right hand forehand swing tennis serve |
| **gesture**                   |             | right arm swipe to the right right arm swipe to the left right hand wave right hand draw x right hand draw circle (clockwise) right hand draw circle (counter clockwise) draw triangle |
| **daily action**              |             | arm curl two hand front clap right arm throw cross arms in the chest two hand push knock on door catch an object pick up and throw |
| **Inertial sensor on right thigh** | training practice | jogging in place walking in place sit to stand stand to sit forward lunge squat |
2.2. Feature extraction
Based on previous inertial sensors for HAR research, we used the sliding window to divide the sequence into 8 parts and the half of the sliding window as a sliding step, then the sampled signal data were processed. The robust features come from different kind of statistical analysis methods in each sliding window and the whole action sequence.

The mean of a window is obtained as:

$$\bar{w} = \frac{1}{N} \sum_{i=1}^{N} w_i$$  \hspace{1cm} (1)

The standard deviation of a sliding window can be obtained as:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (w_i - \bar{w})^2}$$  \hspace{1cm} (2)

The highest value in a fixed-length sliding window is determined as:

$$m = \max(w)$$  \hspace{1cm} (3)

The least value in a fixed-length sliding window is determined as:

$$n = \min(w)$$  \hspace{1cm} (4)

The kurtosis value in a sequence of actions is calculated as:

$$K = \frac{\sum_{i=1}^{k} (x_i - \bar{x})^4 f_i}{n s^4}$$  \hspace{1cm} (5)
The skewness value in a sequence of actions is obtained as:

\[ S = \frac{\sum_{i=1}^{k} (x_i - \bar{x})^3}{n\sigma^3} \]  

(7)

The information entropy in a sequence of actions is obtained as:

\[ H(X) = \sum_{i=1}^{K} p(x_i) \log p(x_i) \]  

(8)

Typical feature sets for HAR includes not only the above statistical functions, but also peak-to-peak value, waveform factor, crest factor and pulse factor.

2.3. Dimension reduction

The robust features are extracted from inertial sensor signals with high data dimensions. In order to increase the speed of computing, we used PCA to achieve dimension reduction. In PCA, the raw data will be mapped to a new vector space by the covariance matrix, where less data can be used to replace the source data. The implementation of PCA is as follows:

\[ C = \frac{1}{n} XX^T \]  

(9)

\[ X = (X_1, X_2, \ldots, X_n) \]  

(10)

\[ C \] is the source data, \( C \) is the covariance matrix, then we performed eigenvalue decomposition of covariance matrices \( C \) to obtain the solution of principal component analysis:

\[ W^* = (\omega_1, \omega_2, \ldots, \omega_k) \]  

(11)

The first 78 principal components account for 95% of the total data, which were selected as the identification data. The dimensionality reduction results of PCA are presented in Fig. 3.

![Fig 3. Principal component contribution rate.](image-url)
3. RESULTS
In this experiment, a total of 27 types of actions characterized by inertial sensor signals in the UTD-MHAD data set were classified. After feature extraction and PCA dimensionality reduction, 78-dimensional data were selected as effective features. The whole effective features was randomly divided into a testing (75%) and a training set (25%). Training sets are used to train different machine learning algorithms to build classifier models, such as ELM and SVM, then verify on the test set. Here is a brief introduction of the two classifier:

1) ELM: ELM is a forward propagation neural network, which also consists of input layer, hidden layer and output layer. Compared with the traditional neural network, the operation speed of ELM is more faster.

2) SVM: SVM is the most classic classifier algorithm in traditional machine learning, which uses kernel functions to map features from low-dimensional space to high-dimensional space, and finds a hyperplane for classification. Essentially, SVM is a convex quadratic programming problem.

In the SVM, we used a gaussian radial basis function with a radius of 2.0 as the kernel function, and performed a five-fold cross-validation by the grid search method to obtain the optimal penalty factor $c = 9.1896$ and the kernel parameter $g = 0.0825$. In the ELM, we obtained the optimal classification results by changing the number of hidden neurons. When the number of hidden neurons is 5000, we got the best classification results. The whole classification results were obtained by taking the mean value of 100 experiments. In Tab. 2, we compared the classification accuracy of the above two methods, and the accuracy of classification after feature extraction is about 83%, and the standard deviation is within 3%. The experimental result shows that the proposed method is about 16.3% higher than the method used by Chen.

| Classifier | Overall accuracy |
|------------|------------------|
| ELM        | 83.49±2.33%      |
| SVM        | 83.47±2.23%      |
| Chen et al | 67.2%            |

4. Conclusions And Future Work
The recognition and classification of the human motion are important for prediction of health status and behavioral goals. For the purpose of privacy protection, inertial sensor signals were used to classify 27 different actions in this study. First, in order to remove data noise, we used a smoothing filter to process the data. Then, the sliding window was used to sample action sequence and the statistically analyze used in the sliding window to obtain the maximum, minimum, mean value, peak-to-peak value, information entropy, skewness, kurtosis, waveform factor, crest factor, pulse factor. In order to improve the speed and accuracy of classification, we reduced the dimensionality of the above robust features and selected the top 78 principal components, which can represent 95% information of the source data. Finally, the robust features was randomly divided into a testing (75%) and a training set (25%) and the ELM and SVM were used for motion recognition. The most noticeable results of the accuracy is about 16% higher than the result of Chen. The experiment results indicated that the proposed method is of practical value and the accuracy rate can reach 83.49%.

This study is mainly used for human motion classification based on inertial sensor motion data and has achieved good classification accuracy. However, this study still has certain limitations. The sample set in the data set is not enough to train the ideal classifier. In the future, we plan to focus on more efficient features and real-time motion recognition.
Acknowledgments
The corresponding author of this article is Professor Hui Cao. This work was funded by the National Natural Science Foundation (No. 81973981, No. 81473708).

References
[1] Guerra J, Uddin J, Nilsen D, et al. Capture, learning, and classification of upper extremity movement primitives in healthy controls and stroke patients[C]//2017 International Conference on Rehabilitation Robotics (ICORR). IEEE, 2017: 547-554.
[2] Wang Y, Zhao Y, Chan R H M, et al. Volleyball skill assessment using a single wearable micro inertial measurement unit at wrist [J]. IEEE Access, 2018, 6: 13758-13765.
[3] Mumtaz N, Arif S, Qadeer N, et al. Development of a low cost wireless IMU using MEMS sensors for pedestrian navigation[C]//2017 International Conference on Communication, Computing and Digital Systems (C-CODE). IEEE, 2017: 310-315.
[4] Seel T, Raisch J, Schauer T. IMU-based joint angle measurement for gait analysis [J]. Sensors, 2014, 14 (4): 6891-6909.
[5] Wittmann F, Lamercy O, Gonzenbach R R, et al. Assessment-driven arm therapy at home using an IMU-based virtual reality system[C]//2015 IEEE international conference on rehabilitation robotics (ICORR). IEEE, 2015: 707-712.
[6] Robert-Lachaine X, Mecheri H, Larue C, et al. Validation of inertial measurement units with an optoelectronic system for whole-body motion analysis [J]. Medical & biological engineering & computing, 2017, 55 (4): 609-619.
[7] Tao L, Vidal R. A Discriminative and Interpretable Skeletal Motion Representation for Action Recognition [C]// 2015 IEEE International Conference on Computer Vision Workshop (ICCVW). IEEE, 2015:303-311.
[8] Ofli F , Chaudhry R , Kurillo G , et al. Sequence of the Most Informative Joints (SMIJ): A new representation for human skeletal action recognition [C]// 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops). IEEE Computer Society, 2012:8-13.
[9] Guerra, Jorge, et al. "Capture, Learning, and Classification of Upper Extremity Movement Primitives in Healthy Controls and Stroke Patients." IEEE Conference on Rehabilitation Robotics (ICORR 2017) IEEE, 2017.
[10] Lin B, Wei X, Zhao J. Automatic recognition and classification of multi-channel microseismic waveform based on DCNN and SVM [J]. Computers & Geosciences, 2019, 123: 111-120.
[11] Anguita D , Ghio A , Oneto L , et al. Human Activity Recognition on Smartphones Using a Multiclass Hardware-Friendly Support Vector Machine [J]. 2012.
[12] Chen C, Jafari R, Kehtarnavaz N. UTD-MHAD: A Multimodal Dataset for Human Action Recognition Utilizing a Depth Camera and a Wearable Inertial Sensor//2015 IEEE International Conference on Image Processing (ICIP), 2015:168-172.