

Recognizing Human Daily Activity Using A Single Inertial Sensor

Chun Zhu and Weihua Sheng
School of Electrical and Computer Engineering
Oklahoma State University
Stillwater, OK, 74078, USA
email: chunz, weihua.sheng@okstate.edu

Abstract—As robot assisted living is becoming increasingly important for elderly people, human daily activity recognition is necessary for human-robot interaction. In this paper, we proposed an approach to daily activity recognition for elderly people. This approach uses a single wearable inertial sensor worn on the right thigh of a human subject to collect motion data. This setup can reduce the obtrusiveness to the minimum. Human daily activities can be recognized in two steps. First, two neural networks are used to classify the basic activities. Second, the activity sequence is modeled by an HMM to consider the sequential constraints exhibited in human daily life and the modified short-time Viterbi algorithm is used for realtime daily activity recognition as the fine-grained classification. We conducted experiments in a mock apartment environment and the obtained results proved the effectiveness and accuracy of our approach.

Index Terms—Activity recognition, assisted living, wearable computing.

I. INTRODUCTION

A. Motivation

With the growth of elderly population in the past decade, more seniors live alone as the sole occupant of a private dwelling than any other population group. The society need take more care of elderly people. Therefore, robot assisted living is getting more attention to help elderly people live a better life. We are developing a smart assisted living (SAIL) system [1], [2] to provide support to elderly people in their daily life. Automated recognition of human daily activities is necessary for human-robot interaction (HRI) [3] in assisted living systems.

There are two main methods for daily activity recognition: vision-based [4] and wearable sensor-based [5]. Vision-based systems can observe full human body movements. However, it is very challenging to recognize human activities through images due to the large volume of data and the data association problem for multiple human subjects. In addition, visual data is easy to be influenced by environments, such as poor lighting conditions and occlusion. Compared to vision-based systems, wearable sensor-based systems have no data association problem and also have less data to process, but it is obtrusive to the user if there are many wearable sensors on the human body.

In this paper, we propose an approach to recognizing human daily activities from a single wearable inertial sensor,

which can reduce the obtrusiveness to the minimum. Compared to vision-based systems, motion sensor-based systems can significantly reduce the computational complexity. We consider the sequential constraints in human daily activities in order to recognize daily activities with a single motion sensor.

This paper is organized as follows. The rest of Section I introduces the related work in this area. Section II describes the hardware platform for the proposed human daily activity recognition system. Section III explains the HMM-based activity recognition using the modified short-time Viterbi algorithm. The experimental results are provided in Section IV. Conclusions are given in Section V.

B. Related Work

In recent years, many approaches are developed for human daily activity recognition. Traditional human daily activity recognition is based on visual information, which involves the pattern recognition of the trajectories of different human body parts. More research works can be found in the survey by Moeslunda *et al.* [4].

Since the computational complexity of vision-based system is high, wearable sensor-based activity recognition has been gaining attention. Inertial sensors are widely used to capture human motion data. Compared to vision-based recognition, wearable sensor-based recognition has two advantages. First, for vision-based recognition, cameras need to be installed prior to the experiments and vision data are usually prone to the influence of environmental factors, such as poor lighting conditions and occlusion. On the contrary, wearable sensors will not be affected by surroundings. Second, wearable sensor-based activity recognition requires less data compared to vision-based recognition. Most wearable sensor systems use multiple sensor nodes to capture motion data. For example, Bao *et al.* [6] used five small biaxial accelerometers worn on different parts of the body. Differences in feature values computed from FFTs are used to discriminate between different activities. Sensors of other modalities, such as air pressure sensor, microphones, and temperature sensors can be used to provide complementary information to motion data and detect variety of activities. However, wearable sensor systems are usually obtrusive and inconvenient to the human subject, especially when there are many sensors. On the

other hand, reducing the number of sensors will increase the difficulty of distinguishing the basic daily activities due to the inherited ambiguity. For example, Aminian *et al.* [7] used two inertial sensors strapped on the chest and on the rear of the thigh to measure the chest acceleration in the vertical direction and the thigh acceleration in the forward direction, respectively. They can detect sitting, standing, lying, and dynamic (walking) activities from the direction of the sensors. However, they cannot discriminate different types of the dynamic activities. Najafi *et al.* [5] proposed a method to detect stationary body postures and walking of the elderly using one inertial sensor attached to the chest. Wavelet transform was used in conjunction with a kinematics model to detect different postural transitions and walking periods during daily physical activities. Because this method has no error correction function, a mis-detection of a postural transition will cause accumulative errors in the recognition. In addition, they did not recognize activities in real-time, which is important to robot assisted living.

Many algorithms have been developed over the years in pattern recognition for human daily activities. There are mainly three types of recognition methods: the heuristic analysis methods [7], the discriminative methods [8], the generative methods [9], and some combinations of them [10]. On the other hand, only a few research focused on real-time activity recognition in previous work, which is important for elderly care. For example, Lo *et al.* [11] used a Multivariate Gaussian Bayes classifier to classify different activities (reading, walking and running) in real-time based on a sensor node with multiple modality channels of data. However, they cannot recognize transitional activities.

In this paper, we used an HMM to model the sequential constraints in human daily life and modified the short-time Viterbi algorithm [12] to decode detailed activities from only a single wearable inertial sensor.

II. HARDWARE PLATFORM OVERVIEW

Our proposed hardware system for human daily activity recognition is shown in Figure 1. We use one inertial sensor and a PDA to collect the sensor data and transfer them to the PC server. The sensor is worn on a thigh of the human subject to significantly reduce the obtrusiveness. The prototype of the motion data collection device is shown in Figure 2. The nIMU sensor from MEMSense, LLC [13], which provides 3D acceleration and angular velocity, is connected to an HP iPAQ PDA through RS422/RS232 serial converter and the PDA sends data to a desktop computer through WiFi.

The sensor provides 3D acceleration and 3D angular velocity at a frequency of 150Hz. Since we find that the angular velocity exhibit similar properties as the accelerations when a human subject performs daily activities, we only collect the 3D acceleration as the raw data, which is represented as:

$$V = [a_x, a_y, a_z]^T \quad (1)$$

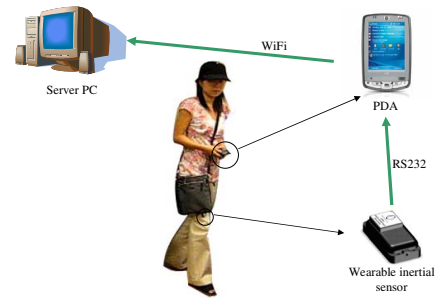


Fig. 1. The overview of the hardware platform for human daily activity recognition.

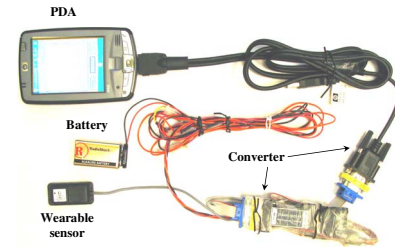


Fig. 2. The inertial sensor data collection prototype.

where a_x , a_y and a_z are the acceleration along direction of x , y and z , respectively.

III. ACTIVITY RECOGNITION USING A SINGLE MOTION SENSOR

Our proposed activity recognition algorithm aims to recognize different daily activities in an indoor environment by using a single wearable sensor. Eight daily activities are to be detected: sitting, standing, lying, walking, sit-to-stand, stand-to-sit, lie-to-sit, and sit-to-lie. The activities can be divided into two kinds: stationary and motional activities. Figure 3 shows the classification of the eight activities into stationary and motional activities. The number to the right of the activity is activity ID.

There are two steps in the recognition algorithm:

- 1) Coarse-grained classification. This step combines the outputs of two neural networks and produces a basic classification.
- 2) Fine-grained classification. This step considers the sequential constraints of human daily activity using an HMM and applies a modified short-time Viterbi algorithm [12] to realize real-time activity recognition in order to generate the detailed activity types.

A. Neural Network-based Coarse-grained Classification

Figure 4 shows the neural network-based coarse-grained classification. Although simply using thresholds on the features can also classify stationary and motional activities, it is required to manually observe the data to set the thresholds. On the contrary, the neural network is a combination of

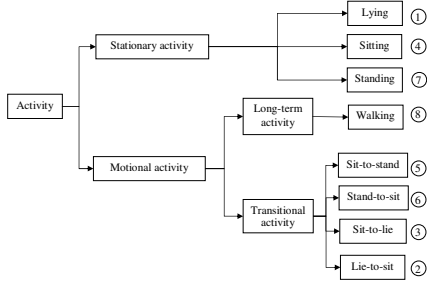


Fig. 3. The taxonomy of human daily activities.

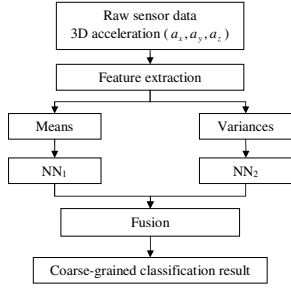


Fig. 4. The neural network-based coarse-grained classification.

multiple thresholds for different features, and can be obtained through training.

1) *Feature Extraction*: In the coarse-grained classification module, feature extraction is applied on the raw sensor data sampled at 150 Hz. We process the raw data using a buffer of 150 data points (1 second). Let D_m represent the m th buffer in realtime processing,

$$D_m = [V_1 \quad V_2 \quad \dots \quad V_{150}]^T \quad (2)$$

The output of feature extraction is F_m , which includes the means and variances of the 3D acceleration.

$$F_m = [\mu_m \quad \sigma_m^2]^T = [\mu_x \quad \mu_y \quad \mu_z \quad \sigma_x^2 \quad \sigma_y^2 \quad \sigma_z^2]^T \quad (3)$$

where $\mu_m = [\mu_x, \mu_y, \mu_z]^T$, and $\sigma_m^2 = [\sigma_x^2, \sigma_y^2, \sigma_z^2]^T$.

2) *Neural Networks*: Two neural networks NN_1 and NN_2 are applied on μ_m and σ_m^2 , respectively. NN_1 is used to detect the direction of the thigh, which is 0 and 1 for horizontal and vertical, respectively. Both NN_1 and NN_2 have a three-layer structure. Let $T_m^{(1)}$ be the output of NN_1 :

$$T_m^{(1)} = \text{hardlim}(f^2(W_1^2 f^1(W_1^1 \mu_m + b_1^1) + b_1^2) - 0.5) \quad (4)$$

where W_1^1, W_1^2, b_1^1 and b_1^2 , are the parameters of NN_1 , which can be trained through the labeled data. The function f^1 and f^2 in both neural networks are Log-Sigmoid function, so that the performance index of the neural networks is differentiable and the parameters can be trained using the back-propagation method [14].

TABLE I
NEURAL NETWORKS FUSION RULES

NN_2	NN_1	
	horizontal	vertical
stationary	Lying and sitting	Standing
movement	All other types (transitions and walking)	

NN_2 is used to detect the intensiveness of the motion of the thigh, which is 0 and 1 for stationary and movement, respectively. Let $T_m^{(2)}$ be the output of NN_2 :

$$T_m^{(2)} = \text{hardlim}(f^2(W_2^2 f^1(W_2^1 \mu_m + b_2^1) + b_2^2) - 0.5) \quad (5)$$

where W_2^1, W_2^2, b_2^1 and b_2^2 , are the parameters of NN_2 , which can also be trained.

3) *Fusion of the Output of Neural Networks*: A fusion function integrates $T^{(1)}$ and $T^{(2)}$ and produces O as the coarse-grained classification. The output of the neural network fusion is: (1) $O \in A_m$ iff $T^{(2)} = 1$ (NN_2 outputs movement): walking and transitional activities; (2) $O \in A_{hs}$ iff $T^{(1)} = 0$ and $T^{(2)} = 0$ (NN_1 outputs horizontal and NN_2 outputs stationary): lying and sitting. (3) $O \in A_{vs}$ iff $T^{(1)} = 1$ and $T^{(2)} = 0$ (NN_1 outputs vertical and NN_2 outputs stationary): standing. The fusion rules are shown in Table I.

B. HMM-based Fine-grained Classification

Due to the inherited ambiguity, It is hard to distinguish the detailed activities from the result of the coarse-grained classification. Some prior knowledge can be used to help model the sequential constraints. Because human daily activities usually exhibit certain sequential constraints, the next activity is highly related with the current activity. Therefore, we can utilize this sequential constraint to distinguish the detained activities. We use a first-order HMM to model such constraints and solve it using a modified short-time Viterbi algorithm. In this paper, we focus on human daily activity recognition for the elderly, which assume that the human subject moves slowly and does not exhibit intensive activities for long periods of time.

1) *Hidden Markov Model for Sequential Activity Constraints*: We assume that the human subject always have a stationary activity for a short time to segment the activities, which is usually true for elderly people. For example, the human subject rises from the chair, stands for a short time, and then starts walking. The standing activity separates the two motional activities. The sequential constraints in fine-grained classification step are referred to as the transitions between different activities. Let S_i be the i^{th} activity in a sequence. S_i depends on its previous activity S_{i-1} and will decide its following activity S_{i+1} in a probabilistic sense. Therefore, we model the activity sequence using an HMM.

An HMM can be used for sequential data recognition. It has been widely used in speech recognition, handwriting

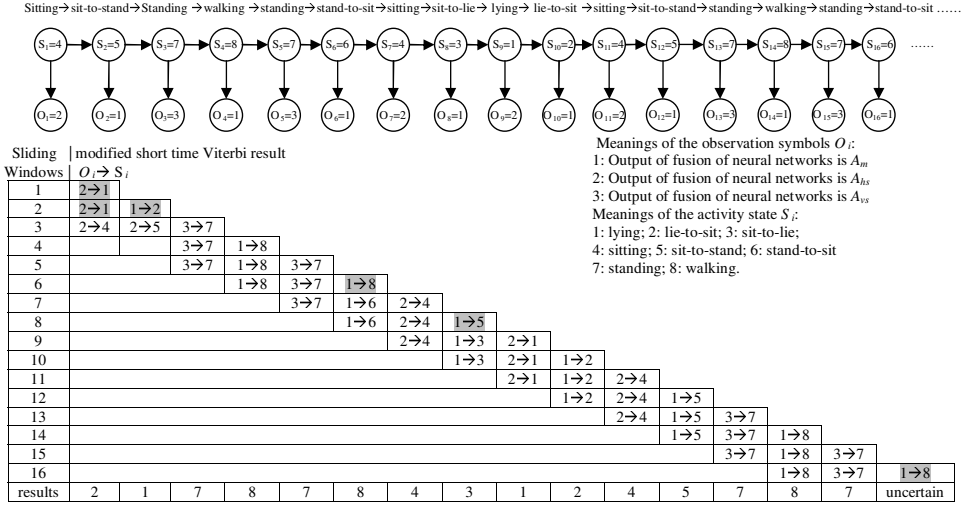


Fig. 5. An sample of activity sequence decoded by the modified short-time Viterbi for HMM.

recognition, and pattern recognition [9]. HMMs can be applied to represent the statistical behavior of an observable symbol sequence in terms of a network of states. An HMM is characterized by a set of parameters $\lambda = (M, N, A, B, \pi)$, where M, N, A, B , and π are the number of distinct states, the number of discrete observation symbols, the state transition probability distribution, the observation symbol probability distributions in each state, and the initial state distribution, respectively. Generally $\lambda = (A, B, \pi)$ is used to represent an HMM with a pre-determined size.

In our implementation, the HMM has eight different states ($M = 8$), which represent eight different activities, and three discrete observation symbols ($N = 3$), which stand for three distinct outputs O_i (A_{hs} , A_{vs} , and A_m) of the coarse-grained classification module. The parameters of the HMM can be trained by observing the activity sequence of the human subject for a period of time. The top part of Figure 5 shows an example of the activity sequence, where each circled S_i is the activity state and O_i is the observed symbol obtained through the fusion of the two neural networks.

2) *Online State Inference Using Short-time Viterbi Algorithm*: Since the standard Viterbi algorithm can only deal with offline process for HMM, we modified the short-time Viterbi algorithm [12] to recover the detailed activity types. Figure 6 shows the decoding problem. We obtain the observation O_i from the coarse-grained classification step. In the fine-grained classification step, the detailed types need to be decoded, which is a mapping from one of three distinct observation values to one of the eight activities.

For the standard Viterbi algorithm [15], the problem is to find the best state sequence when given the observation sequence $O = \{O_1, O_2, \dots, O_n\}$ and the HMM parameters $\lambda = (A, B, \pi)$. In order to choose a corresponding state sequence which is optimal in some meaningful sense, the

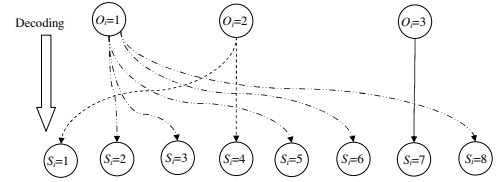


Fig. 6. The decoding of activities.

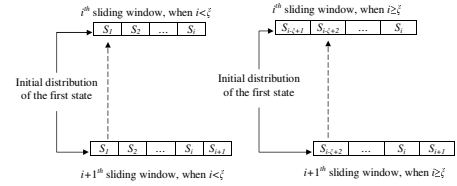


Fig. 7. The initial state corresponding to different sliding windows.

standard Viterbi algorithm considers the whole observation sequence, which does not fit for real-time implementation. Therefore, we propose the modified short-time Viterbi algorithm for online daily activity decoding.

Let $W(i, \xi)$ be the i^{th} sliding window on the observation sequence, where ξ ($\xi \geq 3$) is the length of the sliding window.

$$W(i, \xi) = \begin{cases} \{O_1, O_2, \dots, O_i\}, & (i < \xi) \\ \{O_{i-\xi+1}, O_{i-\xi+2}, \dots, O_i\}, & (i \geq \xi) \end{cases} \quad (6)$$

The result from the short-time Viterbi algorithm is $U(i, \xi)$:

$$U(i, \xi) = \begin{cases} \{S_1, S_2, \dots, S_i\}, & (i < \xi) \\ \{S_{i-\xi+1}, S_{i-\xi+2}, \dots, S_i\}, & (i \geq \xi) \end{cases} \quad (7)$$

$$= \max_{U(i, \xi)} p[U(i, \xi) | W(i, \xi), \lambda] \quad (8)$$

In our approach, the initial state distribution is modified and updated with the result of the previous sliding window.

In the training phase, first we assume uniform distribution and perform recognition using short-time Viterbi algorithm. Second, we summarize the accuracy matrix Ψ for each type of activity, in which each row is used to update the π_i corresponding to the previous result in the testing phase.

Algorithm 1 shows the details for the modified short-time Viterbi algorithm in testing phase. In the testing phase, we use the uniform distribution for π_0 . As the sliding window moves along the observations, the last observation O_i corresponds to the newest activity, which has greater uncertainty for $O_i = A_m$. The state sequence is estimated under the sequential constraints, and except the newest observation in the sequence, other observations can reflect the constraints with the posterior observations. Therefore, we are more confident on the estimation of the previous activities and the initial state distribution π_i is not a constant matrix, which will update with the estimated state sequence for the next sliding window. π_i is probability of the first activity in the $i + 1^{th}$ sliding window, which is the second activity in the i^{th} sliding window. We use the accuracy matrix Ψ to represent the initial probability distribution, which can be learned in the training phase. Figure 7 shows how to find the initial state from the previous sliding window. We update π_i by the following equation:

$$\pi_i(j) = \Psi_{qj} \begin{cases} q = S_1, (i < \xi) \\ q = S_{i-\xi+2}, (i \geq \xi) \end{cases} \quad (9)$$

Algorithm 1 Modified short-time Viterbi for fine-grained classification

```

Initial  $\pi_0, i = 1;$ 
for each new observation  $O_i$  do
  obtain  $W(i, \xi);$ 
  output  $U(i, \xi)$  using Viterbi algorithm based on  $\pi_{i-1};$ 
  MATLAB code, where  $A$  and  $B$  are the parameters of
  HMM,  $o = W(i, \xi); p = \pi_{i-1}; s = U(i, \xi);$ 
   $temp = multinomial\_prob(o, B);$ 
   $s = viterbi\_path(p, A, temp);$ 
  update  $\pi_i$  from Eq 9;
   $i = i + 1;$ 
end for

```

We use the example in Figure 5 to illustrate the on-line state inference using the modified short-time Viterbi algorithm. The human subject made the following activities $S = \{4, 5, 7, 8, 7, 6, 4, 3, 1, 2, 4, 5, 7, 8, 7, 6, \dots\}$. The coarse-grained classification provides the observation symbols $O = \{2, 1, 3, 1, 3, 1, 2, 1, 2, 1, 2, 1, 3, 1, 3, 1, \dots\}$. Each result from the modified short-time Viterbi indicates the mapping from the observation symbols to the detailed activity types. In the result of each sliding window, the newest activity has more uncertainty, especially when $O_i = 1$ for A_m , because the decoding mapping has more candidates. In the gray areas, the

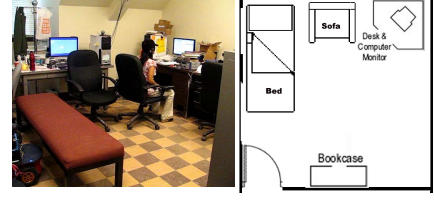


Fig. 8. The layout of the mock apartment.

short-time Viterbi algorithm produces wrong estimates for the newest state in the first sliding window, which are corrected in the following sliding window. In the sliding windows 1 and 2 (row 1 and 2 in the table), because there is too little sequential information, the correct decoded value may not be obtained. As the sequence gets longer (starting from row 3), the detailed activity can be decoded.

IV. EXPERIMENTAL RESULTS

A. Environment Setup

We setup a mock apartment in a lab environment with a dimension of 13.5×15.8 square feet. The layout of the mock apartment is shown in Figure 8. The human subject wore the sensor on the right thigh as shown in Figure 1. Regular daily activities were performed: standing, sitting, sleeping, and transitional activities. Each data set had a duration of about 6 minutes. We recorded video as the ground truth to evaluate the recognition results.

B. Evaluation of the Activity Recognition

In the experiment, we have an output decision value for each second. On the server PC, we use a screen capture software to record the figures which show the output of the recognition results, and compare it with the labeled ground truth recorded from a camera.

Figure 9 shows the result from one set of experiment in the mock apartment. In Figure 9(a), the 3-D acceleration from the sensor indicates stationary and motional activities. Figure 9(b) shows the coarse-grained classification obtained from fusion of the neural networks. Figure 9(c) shows the processing of the modified short-time Viterbi algorithm. The preliminary result is the item on the right edge of each sliding window, which has more uncertainty when the observation value $O_i = 1$. The updated result is the item in the middle of each sliding window, which overlaps the preliminary result of the previous window and can correct the previous mis-classification. In this example, the shadow areas in Figure 9(c) mean that the modified short-time Viterbi algorithm can find correct classifications from the limited observations.

The accuracy in terms of the percentage of correct decisions is listed in Tables II. The values in bold are the percentages of the correct classifications corresponding to the

