ARTICLE IN PRESS

Pervasive and Mobile Computing ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Pervasive and Mobile Computing

journal homepage: www.elsevier.com/locate/pmc

# Motion- and location-based online human daily activity recognition

Chun Zhu, Weihua Sheng *

*School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK, 74078, USA*

ARTICLE INFO

*Article history:*
Received 15 November 2009
Received in revised form 21 September 2010
Accepted 19 November 2010
Available online xxxx

*Keywords:*
Wearable computing
Activity recognition
Sensor fusion

ABSTRACT

In this paper, we proposed an approach to indoor human daily activity recognition which combines motion data and location information. One inertial sensor is worn on the right thigh of a human subject to provide motion data, while an optical motion capture system is used to provide the human location information. Such a combination has the advantage of significantly reducing the obtrusiveness to the human subject at a moderate cost of vision processing, while maintaining a high accuracy of recognition. First, a two-step algorithm is proposed to recognize the activity based on motion data only. In the coarse-grained classification, two neural networks are used to classify the basic activities. In the fine-grained classification, the sequence of activities is modeled by an HMM to consider the sequential constraints. The modified short-time Viterbi algorithm is used for real-time daily activity recognition. Second, to fuse the motion data with the location information, Bayes' theorem is used to update the activities recognized from the motion data. We conducted experiments in a mock apartment and the obtained results proved the effectiveness and accuracy of our algorithms.



## 1. Introduction

### 1.1. Motivation

The past decade has seen a steady growth of elderly population. As the baby boomers comprise nearly 20% of the US population [1], they may bring an increased burden on the medical industry in the near future. Compared to the rest of the population, more seniors live alone as the sole occupants of a private dwelling than any other population group. Therefore, helping seniors live a better life is very important and has great societal benefits. We are developing a smart assisted living (SAIL) system [2,3], which can help elderly people in their daily life. In such assisted living systems, automated recognition of human daily activities is important, which can be used to study behavior related diseases and detect abnormal behaviors. Activity recognition is also indispensable for human–robot interaction (HRI) [4] in assisted living systems where a robot companion can understand the human's intentions through his/her behaviors.

Researchers have developed two main approaches for daily activity recognition: vision-based [5] and wearable sensor-based [6,7]. Vision-based systems can observe full human body movement. However, it is very challenging to recognize human activities through images due to the inherited data association problem and the handling of large volumes of data. Compared to vision-based systems, wearable sensor-based systems have no data association problem and also have less data to process, but it is uncomfortable and obtrusive to the user if there are many wearable sensors on the human body.

In our previous work [2,3], two wired inertial sensors were used to collect motion data and a PDA was required in the system to transfer the data. The activity recognition algorithms only processed the motion data from the sensors. In this paper, we propose an approach that combines motion data from a single wearable inertial sensor and location information

* Corresponding author. Tel.: +1 405 7447590; fax: +1 405 7449198.
*E-mail address:* weihua.sheng@okstate.edu (W. Sheng).

1574-1192/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.pmcj.2010.11.004

Please cite this article in press as: C. Zhu, W. Sheng, Motion- and location-based online human daily activity recognition, Pervasive and Mobile Computing (2010), doi:10.1016/j.pmcj.2010.11.004

from an optical motion capture system to recognize human daily activities. This approach has the following advantages: first, a single wireless inertial sensor worn by the user for motion data collection can reduce obtrusiveness to the minimum; second, less data is required for activity recognition so that the computational complexity is significantly reduced compared to a pure vision-based system; third, the recognition accuracy can be maintained through the fusion of motion and location data.

There are some existing methods using both motion data and image data [8,9]. They used the image data as the major source for activity recognition and added the motion data from an accelerometer as a complemental source to avoid some shortcomings of the vision-based recognition method. Therefore, their algorithms still used complicated image processing. In our approach, since we have found that human daily activities and locations are highly correlated in indoor environments, activities have different probabilities at different locations. Activity recognition from the motion data can be combined with the location data, so that the accuracy of activity recognition can be improved. We only use cameras to obtain the location information and image processing is not required in our activity recognition algorithm.

This paper is organized as follows. The rest of Section 1 introduces the related work in this area. Section 2 describes the hardware platform for the proposed human daily activity recognition system. Section 3 explains activity recognition using motion data only. Section 4 presents the fusion of motion data and location information to improve the recognition accuracy. The experimental results are provided in Section 5. Conclusions and future work are given in Section 6.

### 1.2. Related work

Researchers have made significant progress in the area of human daily activity recognition in recent years. Traditional human daily activity recognition is based on visual information. A typical approach for vision-based recognition has two steps: feature extraction and pattern recognition. In the feature extraction step, activities are analyzed in terms of the tracks of moving body parts, and features are extracted from each image frame [10]. In the pattern recognition step, activities are analyzed using context information of the body parts, which is represented by the extracted features [11]. For a detailed survey of vision-based recognition, please see [5]. However, vision-based activity recognition incurs a significant amount of computational cost, and vision data are usually prone to the influence of environmental factors, such as poor lighting conditions and occlusion.

#### 1.2.1. Wearable sensor-based activity recognition

Due to the advancement in MEMS and VLSI technologies, wearable sensor-based activity recognition has been gaining attention. Inertial sensors are widely used to capture human motion data. For example, Bao et al. [12] used five small biaxial accelerometers worn on different body parts. Differences in feature values computed from FFTs were used to discriminate different activities. The data processing of the five 2-axis accelerometers required significant computational power. Yang et al. [13] built a wireless body sensor system with seven distributed sensor nodes attached to the human body. They obtained high accuracy but the sensor set was power consuming and not convenient for the human subject. Sensors of other modalities can be used to provide complementary information to motion data and detect various activities. For example, Atallah et al. [14] investigated the use of an ear worn activity recognition device combined with wireless ambient sensors for identifying common activities of daily living. Multiple ambient sensors were installed such as door sensors, scales, bed usage sensors, etc. They considered the ambient sensors as other channels of sensing input and the recognition results rely mostly on them. Amft et al. [15] used force sensitive resistors and fabric stretch sensors to detect the contraction of arm muscles and showed that the sensors could provide important information for activity recognition. However, these sensors were obtrusive since they had to be attached to the skin of the human subject.

From the above examples, we can use see that wearable sensor systems are usually obtrusive and inconvenient to the human subject, especially when there are many wearable sensors. However, 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. [16] 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 could detect sitting, standing, lying, and dynamic (walking) activities from the direction of the sensors. However, they could not discriminate different types of the dynamic activities. Najafi et al. [6] 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 did not have any error correction function, a mis-detection of a postural transition would cause accumulative errors in the recognition. In addition, they could not recognize activities in real-time.

#### 1.2.2. Location related activity recognition

There are some existing works using location information to assist human activity recognition. Location is obtained in many different ways. For example, Raj et al. [17] collected GPS data in the outdoor environment and fused it with the measurement from a wearable sensor board. They considered the location information as another parallel data channel in the Bayesian network of activity recognition. However, they could not get detailed indoor location information. We focus on the indoor environment and consider the location information after the activity recognition from a single wearable sensor.

Please cite this article in press as: C. Zhu, W. Sheng, Motion- and location-based online human daily activity recognition, Pervasive and Mobile Computing (2010), doi:10.1016/j.pmcj.2010.11.004

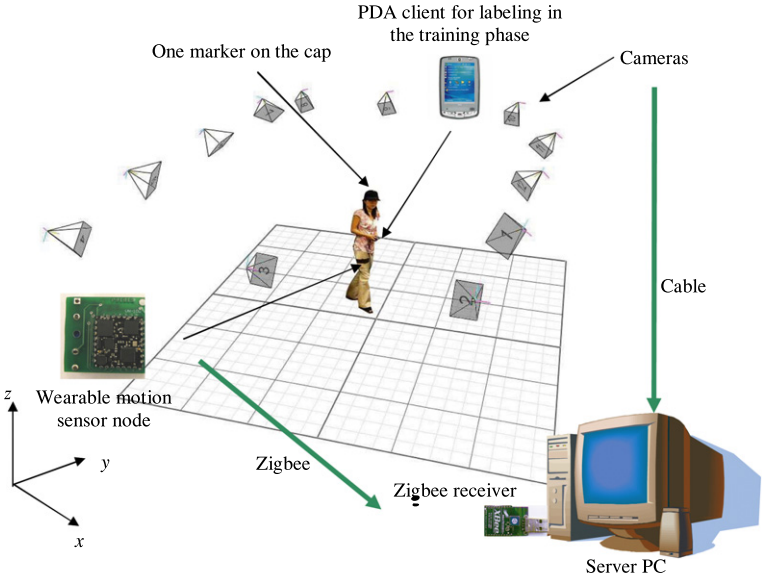



**Fig. 1.** The overview of the hardware platform for human daily activity recognition.

Ince et al. [18] used static home sensors such as door sensors, pressure sensors, and contact sensors to detect the location of a human subject. However, their method had data assignment problems, which had difficulty in distinguishing multiple human subjects. Bieber et al. [9] used a webcam to track the moving human subject. The main objective of the optical analysis was to detect important events. A series of image processing algorithms were applied to extract useful information. In our approach, in order to avoid complicated image processing, an IR source-based 3D camera system is used to obtain the indoor location data, which does not require image processing for human tracking and can also distinguish multiple human subjects.

#### *1.2.3. General algorithms for daily activity recognition*

To recognize human daily activities, many solutions have been developed over the years, including heuristic analysis methods [16], discriminative methods [19,20], generative methods [21], and some combinations of them [22]. Heuristic analysis methods require intuitive analysis on the raw sensor data or the features from data, and the characteristics may be different for individual human subjects. Therefore, it is difficult to find a universal approach for different problems. Discriminative methods and generative methods are both machine learning algorithms and the parameters can be trained using data from different human subjects. However, their disadvantage is the high computational cost. The combination of different methods can achieve better performance than any single one.

## 2. Hardware platform overview

Our proposed hardware system for human daily activity recognition is shown in Fig. 1. We use one inertial sensor to collect the motion data and transfer it to the server. The cameras in the optical motion capture system are used to provide location information. The wearable inertial sensor is synchronized with the video data from the optical system. Thus, the minimum setup of the wearable sensor system is combined with the optical system to facilitate human daily activity recognition. The single sensor setup significantly reduces the obtrusiveness to the human subject. The optical system provides real-time location coordinates of the human subject rather than raw video data, which greatly reduces the computational complexity.

### *2.1. Hardware setup for motion data collection*

Fig. 2 shows the prototype of the new motion sensor node developed using a commercial VN-100 [23]. The node sends data through Zigbee to a receiver on the PC for processing. ZigBee is one type of short-distance wireless communication standard, which is used for a wireless networking range of 10–20 m mainly in homes and offices. Each motion sensor node has an ID to be distinguished from others. Therefore, multi-person activity can also be tracked in this system. Zigbee consumes less power compared to WiFi, but the networking range is limited. Due to the limited size of our lab, we tested our approach using Zigbee. Our method can also work in a larger space when other communication standards are considered, such as WiFi. A PDA is used to label the activities in the training phase but it is not necessary in the real-time testing phase.

Please cite this article in press as: C. Zhu, W. Sheng, Motion- and location-based online human daily activity recognition, Pervasive and Mobile Computing (2010), doi:10.1016/j.pmcj.2010.11.004

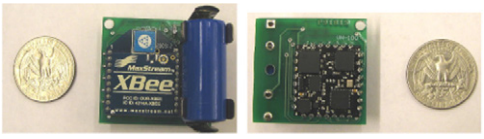



**Fig. 2.** The inertial sensor data collection prototype.

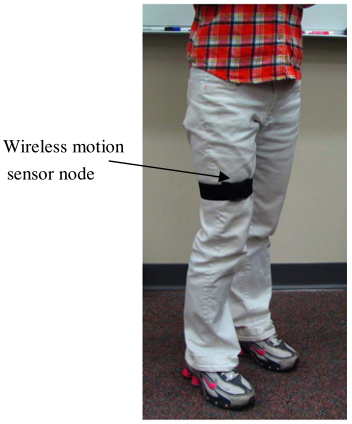



**Fig. 3.** The wireless sensor module is worn on the thigh of the human subject.

Since the position of the attached sensor is very important to activity recognition [7], we collected data using the sensor on different parts of the human body and have found that the thigh is the best location for activity recognition using a single sensor as shown in Fig. 3. The wearable motion sensor node samples the 3D acceleration at a rate of 20 Hz. In the experiments, since normal daily activities are performed following the style of an elderly person, it is observed that the angular velocity exhibits similar properties as the acceleration, we only collect the 3D acceleration as the raw data, which is represented as:

$$V = [a_x, a_y, a_z] \tag{2.1}$$

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

### 2.2. Hardware setup for location information collection

The OptiTrack motion capture system from NaturalPoint, Inc. [24] is marker-based and consists of twelve cameras. The tracking software runs on a PC server to calculate the position of the markers in real-time. The 3D location of the markers can be resolved with millimeter accuracy. Increasing the number of cameras can help improve the tracking performance if needed. The real-time data streaming rate is 100 fps. We downsample the video data to synchronize the inertial sensor data with the video data.

We use one marker attached to a baseball cap to track the human subject. The output coordinate in the 2D ($x$-$y$) space gives us the location information of the human subject, which can be represented as:

$$P = [x, y] \tag{2.2}$$

In real applications, we can use regular cameras instead of the OptiTrack system to calculate the location information, which has much less computational cost compared to activity recognition from raw visual data.

## 3. Activity recognition using a single motion sensor node

We first develop a single wearable sensor-based activity recognition algorithm without considering the location information. Eight daily activities are to be recognized: 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 motion activities. Fig. 4 shows the classification of the eight activities into stationary and motion activities. The number to the right of the activity is the activity ID.

There are two steps in our proposed recognition algorithm: (1) coarse-grained classification and (2) fine-grained classification. The coarse-grained classification step combines the outputs of two neural networks and produces a basic classification. The fine-grained classification step applies a modified short-time Viterbi algorithm [25] to realize real-time activity recognition with the sequential constraints modeled by an HMM, and generates the detailed activity types.

Please cite this article in press as: C. Zhu, W. Sheng, Motion- and location-based online human daily activity recognition, Pervasive and Mobile Computing (2010), doi:10.1016/j.pmcj.2010.11.004

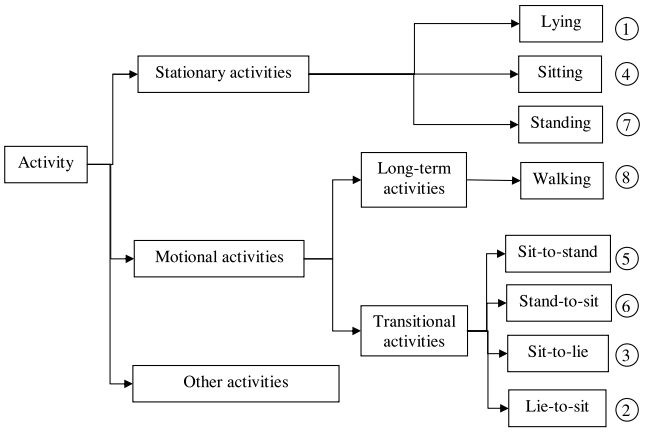



**Fig. 4.** The taxonomy of human daily activities.

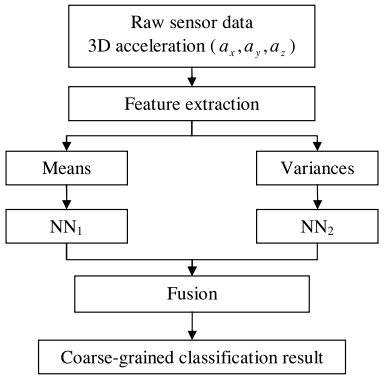



**Fig. 5.** The neural network-based coarse-grained classification.

### 3.1. Neural network-based coarse-grained classification

Fig. 5 shows the neural network-based coarse-grained classification. Although simply using thresholds on the features can also classify stationary and motion activities, it is required to manually observe the data to set the thresholds. On the contrary, the neural network is a combination of multiple thresholds for different features, and can be obtained through training.

#### 3.1.1. Feature extraction

In the coarse-grained classification module, feature extraction is applied on the raw sensor data. We process the raw data using a buffer of 20 data points (1 s). Let $D_m$ represent the $m$th buffer in real-time processing,

$$D_m = \begin{bmatrix} V_1 & V_2 & , \ldots , & V_{20} \end{bmatrix}. \tag{3.3}$$

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

$$\begin{aligned} F_m &= \begin{bmatrix} \mu_m & \sigma_m^2 \end{bmatrix} \\ &= \begin{bmatrix} \mu_x & \mu_y & \mu_z & \sigma_x^2 & \sigma_y^2 & \sigma_z^2 \end{bmatrix} \end{aligned} \tag{3.4}$$

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

#### 3.1.2. Neural networks

Two neural networks $NN_1$ and $NN_2$ are applied on $\mu_m$ and $\sigma_m^2$, respectively. $NN_1$ is used to detect the stationary state 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 $NN_1$:

$$T_m^{(1)} = \text{hardlim}(f^2(W_1^2 f^1(W_1^1 \mu_m + b_1^1) + b_1^2) - 0.5) \tag{3.5}$$

Please cite this article in press as: C. Zhu, W. Sheng, Motion- and location-based online human daily activity recognition, Pervasive and Mobile Computing (2010), doi:10.1016/j.pmcj.2010.11.004

**Table 1**
Neural networks fusion rules.

| $NN_2$ | $NN_1$ | |
|---|---|---|
| | Horizontal | Vertical |
| Stationary | Lying and sitting | Standing |
| Movement | All other types (transitions and walking) | |

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 chosen as a 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 [26].

$NN_2$ 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 $NN_2$:

$$T_m^{(2)} = \text{hardlim}(f^2(W_2^2 f^1(W_2^1 \mu_m + b_2^1) + b_2^2) - 0.5) \tag{3.6}$$

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.

### 3.1.3. Fusion of the output of neural networks

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

## 3.2. 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 to the current activity. Therefore, we can utilize this sequential constraint to distinguish the detailed activities. We use a first order HMM to model such constraints and solve it using a modified short-time Viterbi algorithm.

### 3.2.1. Hidden Markov model for sequential activity constraints

We assume that the human subject always exhibits 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 motion 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 and handwriting recognition [21]. 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 $\pi$ 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.

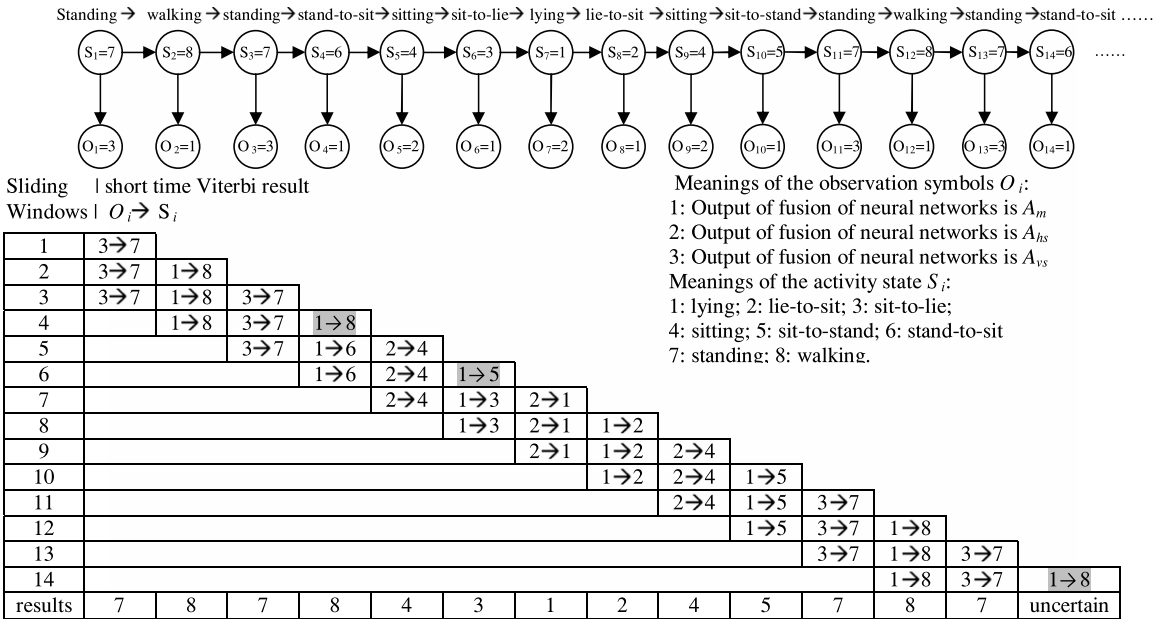
In our implementation, the HMM has eight different states ($M = 8$), which represent eight different activities, and three discrete observation symbols ($N = 3$), which stand for three distinct outputs $O_i$ ($A_{hs}$, $A_{vs}$, and $A_m$) of the coarse-grained classification module. The parameters of the HMM can be trained by observing the activity sequence of the human subject for a period of time. The top part of Fig. 6 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.

### 3.2.2. Online state inference using the short-time Viterbi algorithm

Since the standard Viterbi algorithm can only deal with offline processes for HMM, we modified the short-time Viterbi algorithm [25] to recover the detailed activity types. Fig. 7 shows the decoding problem. The observation $O_i$ is obtained 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 eight activities.

For the standard Viterbi algorithm [27], the problem is to find the best state sequence when given the observation sequence $O = \{O_1, O_2, \ldots, O_n\}$ and the HMM parameters $(A, B, \pi)$. In order to choose a corresponding state sequence which is optimal in some meaningful sense, the standard Viterbi algorithm considers the whole observation sequence, which

Please cite this article in press as: C. Zhu, W. Sheng, Motion- and location-based online human daily activity recognition, Pervasive and Mobile Computing (2010), doi:10.1016/j.pmcj.2010.11.004




| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 3→7 | | | | | | | | | | | | |
| 2 | 3→7 | 1→8 | | | | | | | | | | | |
| 3 | 3→7 | 1→8 | 3→7 | | | | | | | | | | |
| 4 | | 1→8 | 3→7 | 1→8 | | | | | | | | | |
| 5 | | | 3→7 | 1→6 | 2→4 | | | | | | | | |
| 6 | | | | 1→6 | 2→4 | 1→5 | | | | | | | |
| 7 | | | | | 2→4 | 1→3 | 2→1 | | | | | | |
| 8 | | | | | | 1→3 | 2→1 | 1→2 | | | | | |
| 9 | | | | | | | 2→1 | 1→2 | 2→4 | | | | |
| 10 | | | | | | | | 1→2 | 2→4 | 1→5 | | | |
| 11 | | | | | | | | | 2→4 | 1→5 | 3→7 | | |
| 12 | | | | | | | | | | 1→5 | 3→7 | 1→8 | |
| 13 | | | | | | | | | | | 3→7 | 1→8 | 3→7 |
| 14 | | | | | | | | | | | | 1→8 | 3→7 |
| results | 7 | 8 | 7 | 8 | 4 | 3 | 1 | 2 | 4 | 5 | 7 | 8 | 7 | uncertain |

**Fig. 6.** An sample of activity sequence decoded by the short-time Viterbi for HMM.

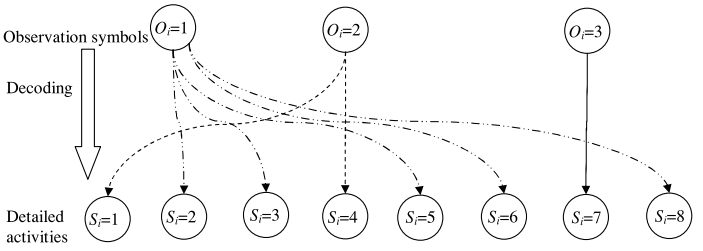



**Fig. 7.** The decoding of activities.

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 $\xi(\xi \geq 3)$ is the length of the sliding window.

$$W(i,\xi) = \begin{cases} \{O_1, O_2, \ldots, O_i\}, & (i < \xi) \\ \{O_{i-\xi+1}, O_{i-\xi+2}, \ldots, O_i\}, & (i \geq \xi). \end{cases} \tag{3.7}$$

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

$$U(i,\xi) = \begin{cases} \{S_1, S_2, \ldots, S_i\}, & (i < \xi) \\ \{S_{i-\xi+1}, S_{i-\xi+2}, \ldots, S_i\}, & (i \geq \xi) \end{cases} \tag{3.8}$$

$$= \arg\max_{U(i,\xi)} p[U(i,\xi)|W(i,\xi), \lambda]. \tag{3.9}$$

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 the short-time Viterbi algorithm. Second, we summarize the accuracy matrix $\Psi$ for each type of activity, in which each row is used to update the $\pi_i$ corresponding to the previous result in the testing phase.

Algorithm 1 shows the details of the modified short-time Viterbi algorithm in testing phase. In the testing phase, we use the uniform distribution for $\pi_0$. As the sliding window moves along the observations, the last observation $O_i$ corresponds to the newest activity, which has greater uncertainty if $O_i$ is $A_m$. 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 $\pi_i$ is not a constant matrix, which will update with the estimated state sequence for the next sliding window. $\pi_i$ is the probability of the first activity in the $i+1$th sliding window, or the second activity in the $i$th sliding window. We

Please cite this article in press as: C. Zhu, W. Sheng, Motion- and location-based online human daily activity recognition, Pervasive and Mobile Computing (2010), doi:10.1016/j.pmcj.2010.11.004

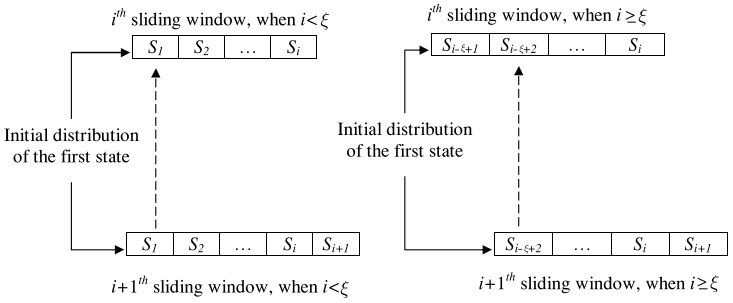



**Fig. 8.** The initial state corresponding to different sliding windows.

**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 $\pi_i$ using Eq. (3.10);
  $i = i + 1$;
**end for**

use the accuracy matrix $\Psi$ to represent the initial probability distribution, which can be learned in the training phase. Fig. 8 shows how to find the initial state from the previous sliding window. We update $\pi_i$ using the following equation:

$$\pi_i(j) = \Psi_{qj} \begin{cases} q = S_1, & (i < \xi) \\ q = S_{i-\xi+2}, & (i \geq \xi). \end{cases} \tag{3.10}$$

We use the example in Fig. 6 to illustrate the modified short-time Viterbi algorithm. The human subject performed the following activities $S = \{7, 8, 7, 6, 4, 3, 1, 2, 4, 5, 7, 8, 7, 6, \ldots\}$. The coarse-grained classification provides the observation symbols $O = \{3, 1, 3, 1, 2, 1, 2, 1, 2, 1, 3, 1, 3, 1, \ldots\}$. Each result from the modified short-time Viterbi indicates the mapping from the observation to the detailed activity types. In the result of each sliding window, the newest activity has more uncertainty, especially when $O_i = 1$ for $A_m$, since the decoding mapping has more candidates. In the gray areas, the short-time Viterbi algorithm produces wrong estimates, which are corrected in the following sliding window.

## 4. Fusion of motion and location data

In indoor environments, human daily activities and locations are highly correlated. Combining the location information and the activity information can improve the accuracy of activity recognition. Given the floor plan of an apartment, we can infer the probability distribution for each specific activity on the 2D map. For example, Fig. 9 shows the probability distribution of "sitting" and Fig. 10 shows the probability distribution of "sit-to-stand" in a typical apartment. In both figures, darker colors indicate higher probability. When the location shows the subject is on the sofa, there is much less probability for "walking". This knowledge can help correct the errors in the single wearable sensor-based activity recognition.

Our overall approach is shown in Fig. 11. Let $\hat{S}_i$ be the $i$th estimated activity from the fine-grained classification step and $L_i$ be the corresponding location from the motion capture system. Bayes' theorem is used to fuse the motion data and the location information to obtain the final results. We utilize a conditional probability distribution function $p(S_i|L_i)$ to represent activity probability distribution given the location information in a layout map. There are two methods to obtain this probability distribution function. First, it can be obtained based on a given floor plan in which locations and activities are correlated using human knowledge. Second, it can be trained by observing the living pattern of a specific human subject for a sustained period of time, which can be more accurate.

We assume that the location measurement is relatively accurate. From Bayes' theorem, the true activity state $S_i$ given the estimated activity $\hat{S}_i$ and the location $L_i$ can be calculated as follows:

$$p(S_i|\hat{S}_i, L_i) \propto p(\hat{S}_i|S_i, L_i)p(S_i|L_i). \tag{4.11}$$

Since we do not consider the location factor in the fine-grained classification step, the activity estimation is independent of the location. Then we have:

$$p(S_i|\hat{S}_i, L_i) \propto p(\hat{S}_i|S_i)p(S_i|L_i) \tag{4.12}$$

Please cite this article in press as: C. Zhu, W. Sheng, Motion- and location-based online human daily activity recognition, Pervasive and Mobile Computing (2010), doi:10.1016/j.pmcj.2010.11.004

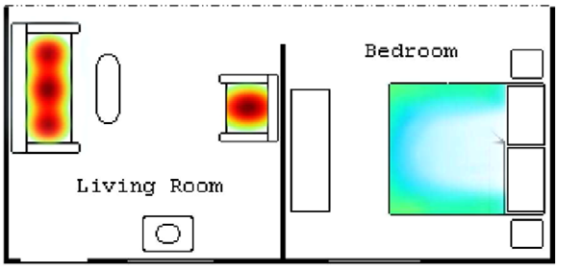



**Fig. 9.** The probability distribution of "sitting" in the map.

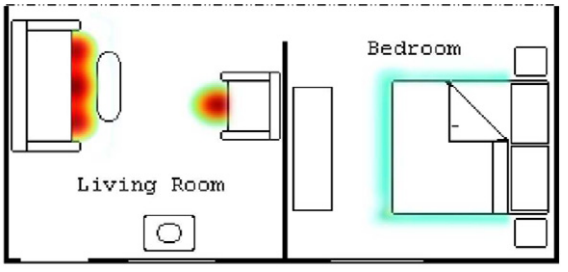



**Fig. 10.** The probability distribution of "sit-to-stand" in the map.

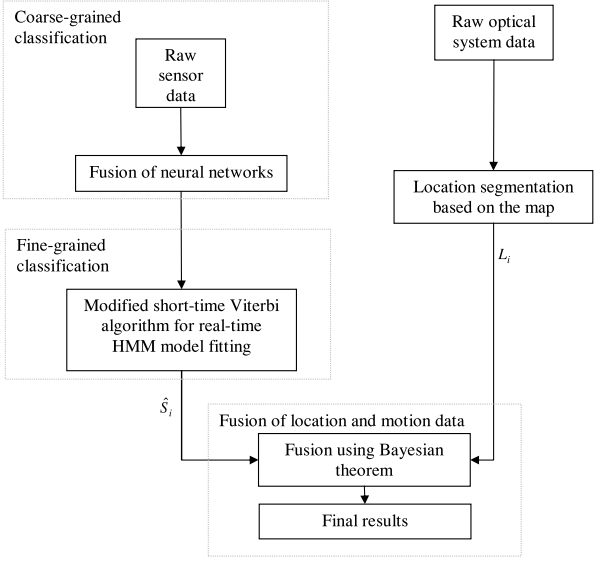



**Fig. 11.** The flow chart of the daily activity recognition algorithm.

where $p(\hat{S}_i|S_i)$ is the probability of observation distribution for each activity. $p(\hat{S}_i|S_i)$ represents the recognition result distribution when the true activity is $S_i$, which can be learned from the accuracy matrix of the fine-grained activity classification.

Finally, the refined activity estimate from the fusion of motion data and location information is obtained as:

$$\hat{S}' = \arg\max_{S_i}(p(S_i|\hat{S}_i, L_i)). \tag{4.13}$$

## 5. Experimental results

### 5.1. Environment setup

We performed the experiments in a mock apartment, which is in a lab environment with a dimension of 13.5 × 15.8 square feet as shown in Fig. 12. The OptiTrack motion capture system is installed on the wall. To simplify the calculation, the given map of the mock apartment is segmented into different areas with corresponding probabilities of activity.

Please cite this article in press as: C. Zhu, W. Sheng, Motion- and location-based online human daily activity recognition, Pervasive and Mobile Computing (2010), doi:10.1016/j.pmcj.2010.11.004

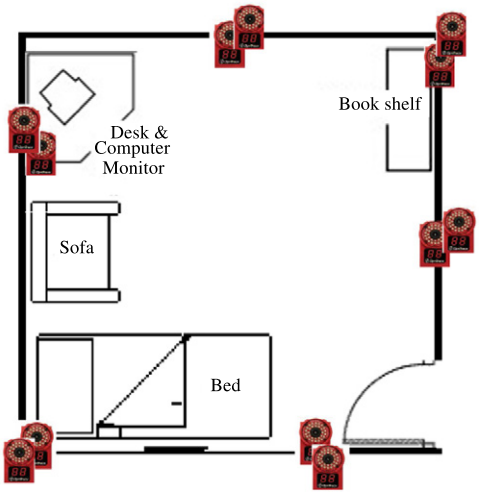



**Fig. 12.** The layout of the mock apartment.

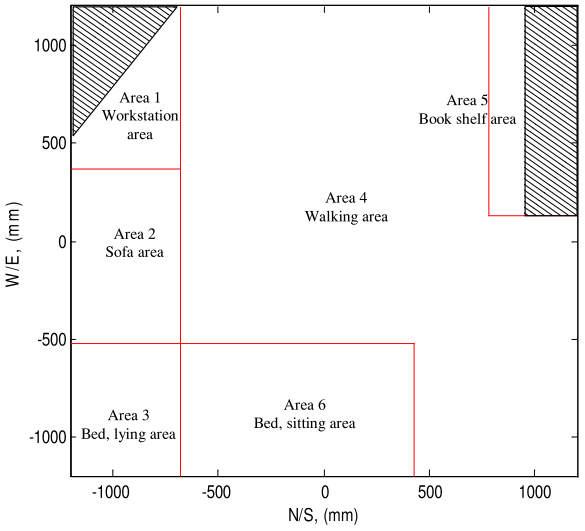



**Fig. 13.** The segmentation of the room.

The coordinates of the human subject given by the OptiTrack system is mapped into $K$ semantic areas $(E_1, E_2, \ldots, E_K)$. The activity distribution given the area $E_q$ can be represented by the conditional probability distribution function $p(S|E_q)$. All locations in the same area have the same activity probability distribution function. According to the furniture layout of the mock apartment and the behavior pattern of the human subject, as shown in Fig. 13, the room is segmented into 6 semantic areas: workstation area, sofa area, bed lying area, bed sitting area, book shelf area and walking area. The behavior pattern of the human subject will affect the segmentations. For example, which side the pillow is on the bed decides "lying" will have higher probability in that side and "sitting" will have higher probability on the other side.

The human subject wore the sensor on the right thigh as shown in Fig. 1. The location of the head was tracked by the OptiTrack system. Regular daily activities were performed: standing, sitting, sleeping, and transitional activities. Each data set had a duration of about 6 min. We recorded video as the ground truth to evaluate the recognition results.

### 5.2. Evaluation of the activity recognition from inertial sensor

In the experiment, we have an output decision value for each second. On the PC server, 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 regular digital camera.

Fig. 14 shows the result from one set of experiments in the mock apartment. In Fig. 14(a), the 3-D acceleration from the sensor indicates stationary and motion activities. Fig. 14(b) shows the coarse-grained classification obtained from the fusion of the two neural networks. Fig. 14(c) shows the processing of the modified short-time Viterbi algorithm. The preliminary result is the number on the right of each sliding window, which has more uncertainty when the observation $O_i$ is 1.

Please cite this article in press as: C. Zhu, W. Sheng, Motion- and location-based online human daily activity recognition, Pervasive and Mobile Computing (2010), doi:10.1016/j.pmcj.2010.11.004

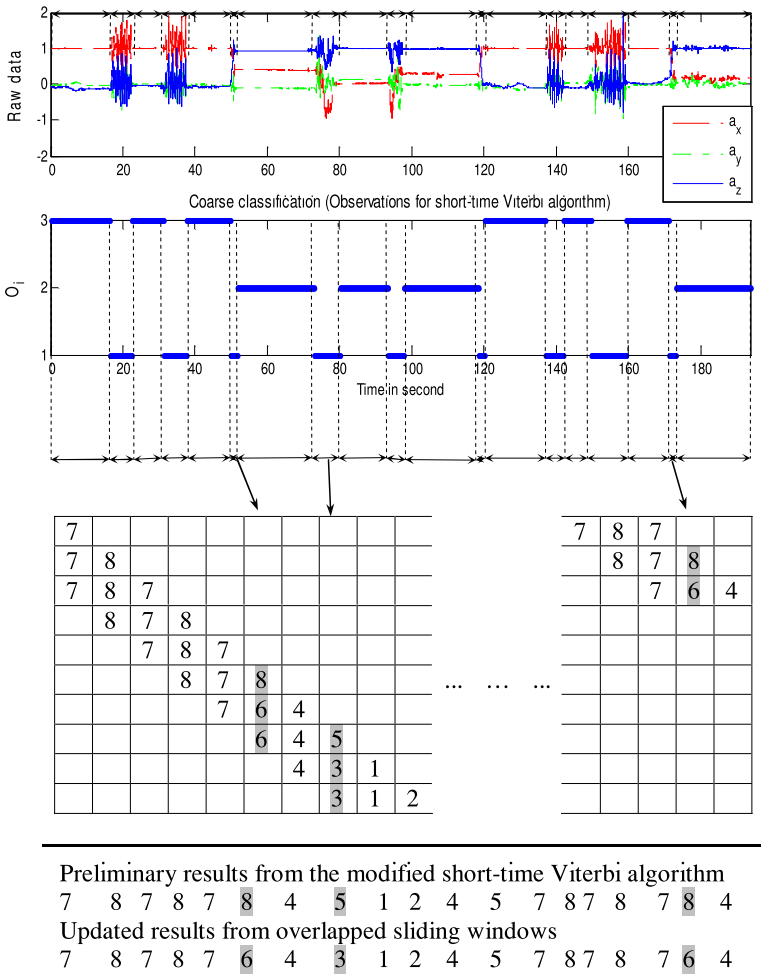



**Fig. 14.** The results of the modified short-time Viterbi algorithm. (a) the 3-D acceleration from the sensor; (b) the coarse-grained classification obtained from fusion of the neural networks; (c) the processing of the modified short-time Viterbi algorithm.

The updated result is the number 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 Fig. 14(c) mean that the modified short-time Viterbi algorithm can find correct classifications from the limited observations.

### *5.3. Evaluation of the fusion of location and motion data*

In the experiment, each output decision value represents the decision for the time window of one second. The accuracy is calculated based on the individual decision made for each sliding window. On the PC server, we use a screen capture software to record the figures which show the output of the recognition results.

The recorded video of the experiment is synchronized with the output of the activity recognition. For the video clips of the experiments, please see [28]. Some significant frames are shown in Fig. 15. From (a) to (j), the top images are from the video of the human subject and the bottom figures are from the PC server screen. In the figure of the recognition results, the top two are the raw sensor data and the segmented location area of the current time step. The middle two are the recognition results from the motion data, and the recognition results from both location and motion data, respectively. The bottom figure is the trace of the human subject obtained from the optical motion capture system. In (a), the human subject goes to area 1 and sits down. In (b), she walks to the bed and sits down. In (c), she lies on the bed. In (d), she sits on the sofa and randomly moves her leg. The results from the motion data are sit-to-lie, and the following activity is lying, which is not correct. The result from the fusion of location and motion data is another transitional activity and the following activity is still sitting, which is correct. Because the random movement of the leg is not one of the pre-defined activities, it will be recognized as one of the closest activities. However the next stationary activity will still be correct because in this area, the probability of sitting is higher than lying. In (e) and (f), she is sitting and moving her leg randomly. Fusion of location can correct the error from lying to sitting. After she stands up from the bed, the results show standing. The previous errors will not accumulate because the decoding mapping of standing has higher confidence so that it will not be affected by the previous errors.

Please cite this article in press as: C. Zhu, W. Sheng, Motion- and location-based online human daily activity recognition, Pervasive and Mobile Computing (2010), doi:10.1016/j.pmcj.2010.11.004

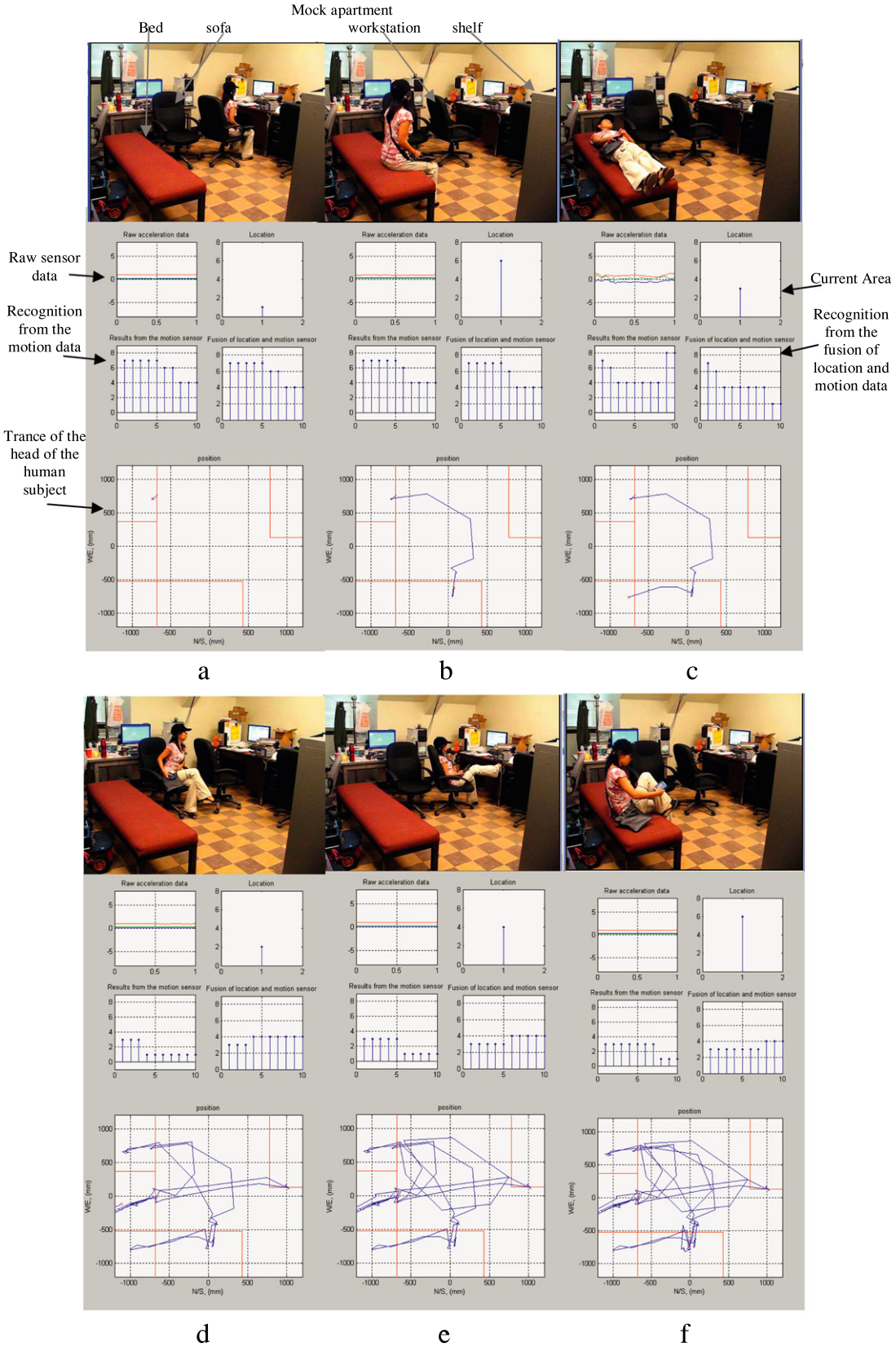



**Fig. 15.** Results captured from video and PC server.

Please cite this article in press as: C. Zhu, W. Sheng, Motion- and location-based online human daily activity recognition, Pervasive and Mobile Computing (2010), doi:10.1016/j.pmcj.2010.11.004

**Table 2**
Decision accuracy obtained from motion data only.

| Test no. | Decision type | | | | | | | | Test accuracy |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | |
| 1 | **0.75** | 0.03 | 0.02 | 0.20 | 0 | 0 | 0 | 0 | **0.75** |
| 2 | 0 | **0.68** | 0.21 | 0 | 0 | 0 | 0 | 0.11 | **0.68** |
| 3 | 0 | 0.25 | **0.67** | 0 | 0 | 0 | 0 | 0.08 | **0.67** |
| 4 | 0.22 | 0 | 0 | **0.78** | 0 | 0 | 0 | 0 | **0.78** |
| 5 | 0 | 0 | 0 | 0 | **0.86** | 0 | 0.05 | 0.09 | **0.86** |
| 6 | 0 | 0 | 0 | 0 | 0 | **0.83** | 0.07 | 0.10 | **0.83** |
| 7 | 0 | 0 | 0 | 0 | 0.05 | 0.03 | **0.92** | 0 | **0.92** |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0.02 | **0.98** | **0.98** |

**Table 3**
Decision accuracy obtained from the fusion of location and motion data.

| Test no. | Decision type | | | | | | | | Test accuracy |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | |
| 1 | **0.90** | 0.03 | 0.02 | 0.05 | 0 | 0 | 0 | 0 | **0.90** |
| 2 | 0 | **0.85** | 0.15 | 0 | 0 | 0 | 0 | 0 | **0.85** |
| 3 | 0 | 0.12 | **0.88** | 0 | 0 | 0 | 0 | 0 | **0.88** |
| 4 | 0.10 | 0 | 0 | **0.90** | 0 | 0 | 0 | 0 | **0.90** |
| 5 | 0 | 0 | 0 | 0 | **0.86** | 0 | 0.05 | 0.09 | **0.85** |
| 6 | 0 | 0 | 0 | 0 | 0 | **0.83** | 0.07 | 0.10 | **0.83** |
| 7 | 0 | 0 | 0 | 0 | 0.05 | 0.03 | **0.92** | 0 | **0.92** |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0.02 | **0.98** | **0.98** |

The accuracy in terms of the percentage of correct decisions of the two methods is listed in Tables 2 and 3. The values in bold are the percentages of the correct classifications corresponding to the specific types of activities. Other numbers indicate the percentages of wrong classifications. Comparing these two tables, the fusion of location and motion data can significantly improve the recognition accuracy compared to the recognition using motion data only. The overall accuracy of our approach is above 0.85, which is higher compared to some recent existing human daily activity recognition methods based on video data only [29–31].

## 6. Conclusions and future work

In this paper, we proposed a method to fuse motion data and location information for human daily activity recognition in an indoor apartment environment. One inertial sensor is worn on the right thigh of the human subject to provide motion data; while an optical motion capture system is used to obtain the location information of the human subject. The activity is first recognized using only the motion data from the inertial sensor by combining the neural networks and the modified short-time Viterbi algorithm. Next, Bayes' theorem is used to integrate the location information to refine the recognition result. Our approach has the advantage of reducing the obtrusiveness and the complexity of vision processing, while maintaining high accuracy of activity recognition. We conducted experiments in a mock apartment environment and the accuracy of the real-time recognition is evaluated. In our future work, we will do experiments on multiple human subjects and larger areas. We will also combine the location and human activities for simultaneous tracking and activity recognition (STAR) [32], which will remove the need of the OptiTrack motion capture system.

## References

[1] Baby boomers aging needs. http://www.Babyboomercaretaker.com, October 2008.
[2] C. Zhu, W. Sheng, Multi-sensor fusion for human daily activity recognition in robot-assisted living, in: Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction, 2009, pp. 303–304.
[3] C. Zhu, W. Sheng, Human daily activity recognition in robot-assisted living using multi-sensor fusion, in: IEEE International Conference on Robotics and Automation, 2009, pp. 2154–2159.
[4] H.A. Yanco, J.L. Drury, Classifying human-robot interaction: an updated taxonomy, in: Proceedings of 2004 IEEE International Conference on Systems, Man and Cybernetics, 2004, pp. 2841–2846.
[5] T.B. Moeslunda, A. Hiltonb, V. Kruger, A survey of advances in vision-based human motion capture and analysis, Computer Vision and Image Understanding (2006) 90–126.
[6] B. Najafi, K. Aminian, A. Paraschiv-Ionescu, F. Loew, C.J. Bula, P. Robert, Ambulatory system for human motion analysis using a kinematic sensor: monitoring of daily physical activity in the elderly, IEEE Transactions on Biomedical Engineering 50 (2003) 711–723.
[7] U. Maurer, A. Smailagic, D.P. Siewiorek, M. Deisher, Activity recognition and monitoring using multiple sensors on different body positions, in: Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks, 2006, pp. 113–116.
[8] G. Bauer, P. Lukowicz, Developing a sub room level indoor location system for wide scale deployment in assisted living systems, Lecture Notes in Computer Science (2008) 1057–1064.
[9] G. Bieber, A. Hoffmeyer, E. Gutzeit, C. Peter, B. Urban, Activity monitoring by fusion of optical and mechanical tracking technologies for user behavior analysis, in: Proceedings of the 2nd International Conference on Pervasive Technologies Related to Assistive Environments, 2009, p. 45.

Please cite this article in press as: C. Zhu, W. Sheng, Motion- and location-based online human daily activity recognition, Pervasive and Mobile Computing (2010), doi:10.1016/j.pmcj.2010.11.004

[10] V. Parameswaran, R. Chellappa, View independent human body pose estimation from a single perspective, Computer Vision and Pattern Recognition (2004).
[11] S. Park, M.M. Trivedi, Multi-person interaction and activity analysis: a synergistic track- and body-level analysis framework, Machine Vision and Applications (2007) 151–166.
[12] L. Bao, S.S. Intille, Activity recognition from user-annotated acceleration data, in: PERVASIVE 2004, 2004, pp. 1–17.
[13] A.Y. Yang, S. Iyengar, S. Sastry, R. Bajcsy, P. Kuryloski, R. Jafari, Distributed segmentation and classification of human actions using a wearable motion sensor network, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2008, pp. 1–8.
[14] L. Atallah, B. Lo, R. Ali, R. King, G. Yang, Real-time activity classification using ambient and wearable sensors, IEEE Transactions on Information Technology in Biomedicine (2009) 1031–1039.
[15] O. Amft, H. Junker, P. Lukowicz, G. Troster, C. Schuster, Sensing muscle activities with body-worn sensors, in: International Workshop on Wearable and Implantable Body Sensor Networks, 2006, p. 4.
[16] K. Aminian, Ph. Robert, E.E. Buchser, B. Rutschmann, D. Hayoz, M. Depairon, Physical activity monitoring based on accelerometry: validation and comparison with video observation, Medical and Biological Engineering and Computing 3 (1999) 304–308.
[17] A. Raj, A. Subramanya, D. Fox, J. Bilmes, Rao-blackwellized particle filters for recognizing activities and spatial context from wearable sensors, Experimental Robotics (2008) 211–221.
[18] N.F. Ince, C.H. Min, A. Tewfik, D. Vanderpool, Detection of early morning daily activities with static home and wearable wireless sensors, EURASIP Journal on Advances in Signal Processing (2008) 1–11.
[19] T. Mitchell, Decision tree learning, Machine Learning (1997) 52–78.
[20] D. Lowd, P. Domingos, Naive Bayes models for probability estimation, in: Proceedings of the 22nd International Conference on Machine Learning, 2005.
[21] L.R. Rabiner, A tutorial on hidden Markov models and selected application in speech recognition, Proceedings of the IEEE 77 (1989) 267–296.
[22] J. Lester, T. Choudhury, N. Kern, G. Borriello, B. Hannaford, A hybrid discriminative/generative approach for modeling human activities, in: Proc. of the International Joint Conference on Artificial Intelligence IJCAI, 2005, pp. 766–772.
[23] VectorNav Technologies, 2010. http://www.vectornav.com/.
[24] Inc. NaturalPoint, OptiTrack™optical motion capture solutions, 2009.
[25] J. Bloit, X. Rodet, Short-time Viterbi for online HMM decoding: evaluation on a real-time phone recognition task, in: Acoustics, Speech and Signal Processing, 2008, ICASSP 2008, IEEE International Conference on, 2008, pp. 2121–2124.
[26] M.T. Hagan, H.B. Demuth, M.H. Beale, Neural Network Design, PWS Publishing Company, 1996.
[27] A.J. Viterbi, Error bounds for convolutional codes and an asymptotically optimal decoding algorithm, IEEE Transactions on Information Theory 13 (1967) 260–269.
[28] C. Zhu, Human daily activity recognition for the assisted living system, 2009. http://www.youtube.com/watch?v=rpQoVUcEeQQ.
[29] O. Brdiczka, P. Reignier, J.L. Crowley, Detecting individual activities from video in a smart home, Lecture Notes in Computer Science (2010) 363–370.
[30] L.N. Abdullah, S.A.M. Noah, Metadata generation process for video action detection, International Symposium on Information Technology, ITSim 2008, 2008, pp. 1–5.
[31] C. Yeo, P. Ahammad, K. Ramchandran, S.S. Sastry, High-speed action recognition and localization in compressed domain videos, IEEE Transactions on Circuits and Systems for Video Technology (2008) 1006–1015.
[32] D. Wilson, C. Atkeson, Simultaneous tracking & activity recognition (star) using many anonymous, binary sensors, in: Proceedings of PERVASIVE, 2005, pp. 62–79.

Please cite this article in press as: C. Zhu, W. Sheng, Motion- and location-based online human daily activity recognition, Pervasive and Mobile Computing (2010), doi:10.1016/j.pmcj.2010.11.004