Human Daily Activity Recognition in Robot-assisted Living Using Multi-sensor Fusion

Chun Zhu and Weihua Sheng

Abstract—In this paper, we propose a human daily activity recognition method by fusing the data from two wearable inertial sensors attached on one foot and the waist of the human subject, respectively. We developed a multi-sensor fusion scheme for activity recognition. First, data from these two sensors are fused for coarse-grained classification in order to determine the type of the activity: zero displacement activity, transitional activity, and strong displacement activity. Second, a fine-grained classification module based on heuristic discrimination or hidden Markov models (HMMs) is applied to further distinguish the activities. We conducted experiments using a prototype wearable sensor system and the obtained results prove the effectiveness and accuracy of our algorithm.

I. INTRODUCTION

A. Motivation

The past decade has seen a steady growth of elderly population. As the baby boomers comprise nearly 20 percent of the U.S. population, which is equal to 76.1 million Americans [1], many of them will turn 65 and are prone to health complications and this may cause an increased burden on the medical industry. 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, elderly people living alone are an at-risk group. Helping them to live a better life is very important and has great societal benefits.

Many researchers are working on new technologies to help elderly people [2], [3]. We are developing a smart assisted living (SAIL) system [4], [5] to help and provide support to elderly people when there is an emergency situation. The SAIL system consists of a body sensor network (BSN) [6], a companion robot, a Smartphone, and a remote health provider. In order to enable natural human-robot interaction, the robot needs to infer the human intentions and situations from the motion data and vital signs of the human subject. For example, when an elderly person falls down accidently, the algorithm will be able to detect this situation and communicate with a companion robot to help the patient. Therefore, there is a great need for the robot to have the capability to recognize the human's activities.

B. Related work

Traditional human activity recognition is through visual information [7], [8]. Recently, due to the advancement in

[1]Chun Zhu is with the School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK, 74078 USA (e-mail: chunz@okstate.edu)

[2]Weihua Sheng is with the School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK, 74078 USA (e-mail: weihua.sheng@okstate.edu)

MEMS and VLSI technologies, wearable sensors-based activity recognition has been gaining attention. Wearable inertial sensors and fiber sensors can be attached to different human body parts to capture kinetic data. Sagawa *et al.* [9] discussed a method to classify human moving behaviors using one acceleration sensor and one air pressure sensor attached on the waist. A slight change of air pressure which results from vertical movement is used to detect the moving styles of going up/down the stairs or in an elevator. Mantyjarvi *et al.* [10] used two sets of accelerometers attached on the left and right side of the hip for activity recognition. Lester *et al.* [11] tested by attaching the sensor board on different parts of the human body and concluded that when training data and testing data are collected from the same locations on the body, the algorithm can pick the right subset of discriminatory features that will work for all those locations. Amft *et al.* [12] investigated the use of force sensitive resistors and fabric stretch sensors that can be easily integrated into clothing. They used these sensors to detect the contractions of arm muscles and argued that the sensors can provide important information for activity recognition.

Recently machine learning algorithms [13] have been widely used for human activity recognition. Many solutions have been developed over the years, including the heuristic analysis methods [14], the discriminative methods [15], [16], the generative methods [17], and some combinations of them. Heuristic analysis methods are through the direct characteristic analysis and the feature description of the data from accelerometers. Aminian *et al.* [14] developed an algorithm based on the analysis of the average and the deviation of the acceleration signal to classify the activities into four categories: lying, sitting, standing and locomotion. Discriminative methods analyze features extracted from sensor data points or segmentations without considering sequential connections in the data. For example, in [10], principal components analysis (PCA) and independent component analysis (ICA) are used in the feature generation process with wavelet transform for the two sets of accelerometers attached on different parts of the human body. Generative methods use generative models for the probability-based observations with hidden parameters. It specifies a joint probability distribution over observation and label sequences, whereas discriminative methods only consider the observed variables, not the sequential data. For example, DeVaul *et al.* [18] developed a two-layer model that combines a multi-component Gaussian mixture model with Markov models to accurately classify a range of user activity states, including sitting, walking, biking, etc. By combining different methods, the advantages of each method can be

better utilized to solve complicated problems. Lester *et al.* [19] presented a hybrid approach to recognize human activities, which combines boosting [20] to discriminatively select useful features and learn an ensemble of static classifiers to recognize different activities, with hidden Markov models (HMMs) [17] to capture the temporal regularities and smoothness of activities. To summarize, heuristic analysis methods require intuitive analysis on the raw sensor data or the features from data, and the characteristics may differ from each individual. Therefore, it is difficult to find a ubiquitous way for observation. On the contrary, since discriminative methods and generative methods are machine learning algorithms, the parameters can be trained using data from different individuals. However, their disadvantage is the high computational cost. The combination of different methods can achieve better performance than any single method.

In this paper, we proposed a 2-step human daily activity recognition method combining the neural networks and the hidden Markov models. In the first step, the fusion of the data from the two wearable sensors generates the coarse-grained classification of human daily activities. In the second step, (1) the heuristic discrimination module or (2) the HMM-based recognition algorithm is used for the fine-grained classification. In this way, the coarse-grained classification controls the direction of the data flow to trigger either the heuristic discrimination module or the HMM-based recognition module in order to save the computation time and enhance the efficiency of the recognition algorithm.

This paper is organized as follows. Section II introduces the overall framework of the human daily activity recognition system. Section III describes the neural network-based activity coarse-grained classification algorithm. Section IV explains the mechanism of the HMM-based activity recognition algorithm. The experimental results are presented in Section V. Conclusions and future work are given in Section VI.
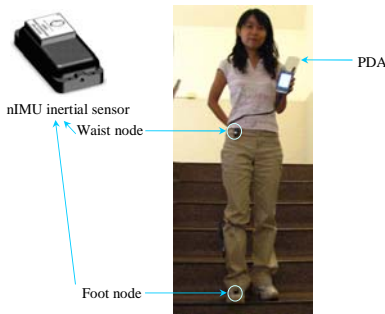

Fig. 1. The prototype of the wearable sensor system for human daily activity recognition.

II. System Overview

The prototype of the wearable sensors for human daily activity recognition is shown in Figure 1. We use two inertial sensors (nIMU NA05-0600F050R) from MEMSense, LLC, which provide 3D acceleration, angular velocity, magnetic data, and temperature. Both inertial sensors are connected to a PDA through RS422/RS232 serial converters, and the PDA sends data to a desktop computer through WiFi, where the data are processed to recognize different activities. The data-collection program for the PDA is written in Visual C++ and the HMM training/recognition program is written in MATLAB.
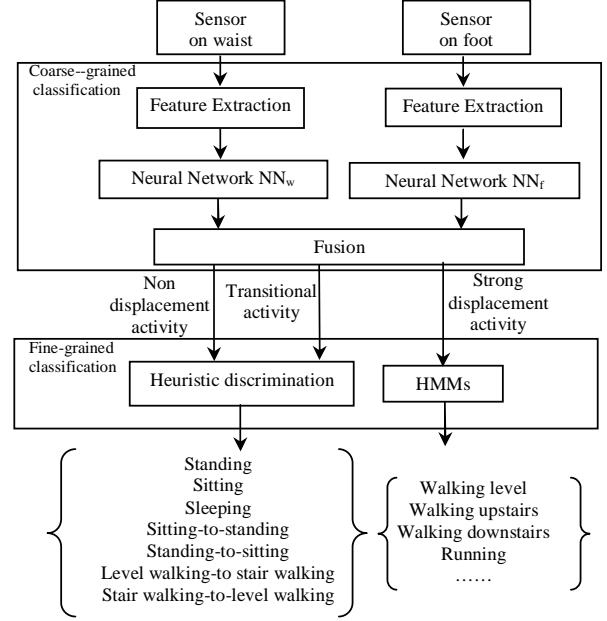

Fig. 2. The overview of the human daily activity recognition system.

The major criteria for our recognition algorithm are the efficiency and accuracy of the recognition algorithm. For embedded computing systems, it is important to design algorithms with light-weight and resource-awareness to save energy and increase the efficiency. In this paper, we propose an innovative method which is self-adaptive and efficient to recognize different human daily activities by two inertial sensors attached to one foot and the waist of the human subject, respectively. Motion data including 3D acceleration and angular velocity are collected, fused and classified. In this project, we consider the following activities: (1) $A_Z =$ **zero displacement activities**: standing, sitting, and sleeping; (2) $A_T =$ **transitional activities**: sitting-to-standing, standing-to-sitting, level walking-to-stair walking, and stair walking-to-level walking; (3) $A_S =$ **strong displacement activities**: walking level, walking upstairs, walking downstairs, and running. More activities can be recognized with extra sensors. For example, cooking and watching TV can be recognized when the environmental audio information is recorded.

Figure 2 shows the block diagram of our algorithm. In the coarse-grained classification module, raw data (acceleration and angular velocity) are processed to obtain the features (mean, variance and covariance of the 3D acceleration and 3D angular velocity), which are fed into the corresponding neural network NN_f and NN_w for foot and waist, respectively. We categorize the outputs of the neural networks NN_f and NN_s into three types: (1) **stationary**, (2) **transitional**, and (3)
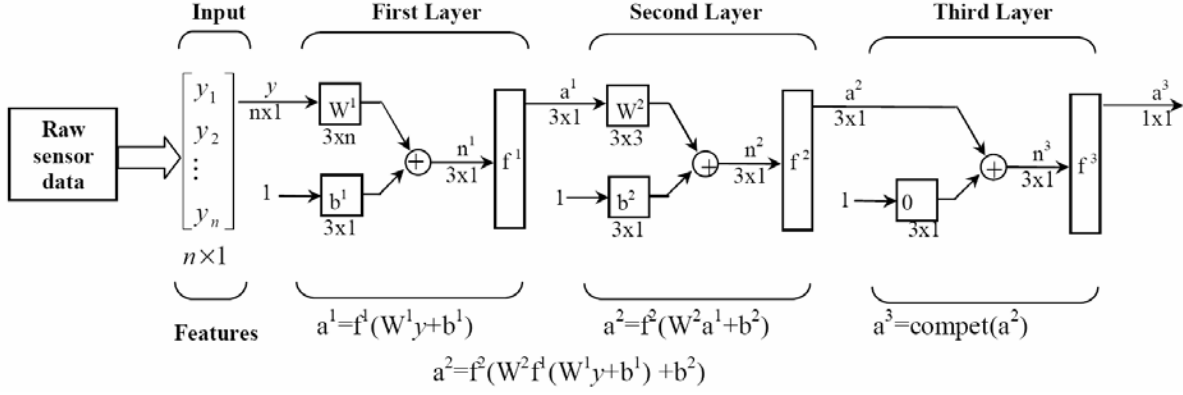
Fig. 3. The structure of 3-layer feed-forward neural network.

cyclic. A fusion module integrates the individual types of foot and waist activities and categorizes the human daily activities according to the following rules in Table I: (1) **zero displacement activities**: $A \in A_Z$ iff $A_w =$ **stationary**; (2) **transitional**: $A \in A_T$ iff ($A_f =$**transitional** and $A_w =$**transitional**) or ($A_f =$ **stationary** and $A_w =$ **transitional**); (3) **strong displacement activities**: $A \in A_S$ iff $A_f =$ **cyclic** and $A_w =$ **cyclic**. All other combinations of foot and waist activities are considered as rare activities and we do not consider them in this project.

In the fine-grained classification module, to further distinguish the stationary activities (such as sitting and standing) and the transitional activities (such as sitting-to-standing and standing-to-sitting), a discrimination module will be applied to consider the previous stationary activity state and decide the type of the current transitional activity. A hidden Markov model (HMM)-based recognition algorithm is applied to further determine the types of the strong displacement activities, which is to recognize the patterns of the continuous time series of data.

TABLE I

Sensor fusion rules

| | | Foot sensor A_f | | |
		Stationary	Transitional	Cyclic
Waist	Stationary	A_Z	A_Z	A_Z
sensor	Transitional	A_T	A_T	–
A_w	Cyclic	–	–	A_S

III. Neural Network-based Segmentation

A. Overview of the neural networks

In this paper, we implemented two feed-forward neural networks [21] to spot non-stationary activities from stationary activities. The neural networks NN_f and NN_w both generate distinctive outputs for stationary activities, transitional and strong displacement activities, respectively. Generally, in daily life, when people stand, sit, stand-to-sit, sit-to-stand, or sleep, their feet do not exhibit extensive motions. Therefore, we use the variance of 3D acceleration and 3D angular velocity to represent the intensity of the movement. The outputs of NN_w

and NN_f are fused and a heuristic segmentation refinement is applied to produce the start and the end point of the segment.

The details of the neural network module is shown in Figure 3: the input is a n-by-1 feature vector extracted from the sensor raw data, which represents n features. f^1 and f^2 are the log-sigmoid functions and f^3 is the 'compet' function. The parameters W^1, b^1, W^2, and b^2 are trained through the back-propagation method. $W^3=1$ and $b^3=0$ are fixed to generate discrete outputs.

B. Training of the neural network

Supervised learning is used to train the neural networks [21]. In the training mode, the experimenter labels the correct type when the human subject is performing daily activities. The label is recorded together with the raw data on the PDA. The back-propagation method is implemented to train the weights and biases of neurons in the first and the second layers. Training starts from a set of random value of weights and biases, and are updated at each iteration to minimize the performance index in order to achieve the minimal of mean square error. However, since not every set of random initial values can ensure that the performance index approaches a certain level, the initial value need to be adjusted in the training step. Moreover, the number of neurons in each layer of the network has to be modified in order to achieve the highest accuracy and avoid over-fit as well.

IV. HMM-based Recognition

In our previous work, we have developed an HMM-based algorithm for hand gesture recognition [4], [5]. For activity recognition, HMMs are applied to the data segmentation from the foot sensor where strong displacement activities are detected by the neural network NN_w and NN_f.

The major steps of the recognition algorithm include data pre-processing, the HMM training phase, the HMM-based recognition phase.

A. Data pre-processing

Data pre-processing is applied to the raw data before they are fed into the hidden Markov models in both the

training phase and the recognition phase. When the computer receives the data, sampled at a rate of 150 Hz from the nIMU sensor, a digital low-pass filter is applied to the 3D acceleration $[a_x, a_y, a_z]^T$ and the 3D angular velocity $[\omega_x, \omega_y, \omega_z]^T$ of the data and produces a 6-component vector $u = [a_x^f, a_y^f, a_z^f, \omega_x^f, \omega_y^f, \omega_z^f]^T$ for each sampling point. Afterward, a sliding-window of 20 points which is about 133 ms in the time domain is used to calculate the average in order to remove the DC components on 3D acceleration and generate the vector $w = [d_x, d_y, d_z]^T$. Because the Fast Fourier Transform (FFT) can give us the power components in the frequency domain, we remove the DC components and offset the means of gravity in order to find the fundamental frequency of the behavior. Since this 3D vector w will be used in the training phase to determine the length for each activity, we propose a new vector including both kinds of information as a result of pre-processing. Finally, a vector of 3D acceleration, 3D angular velocity, and 3D deviation of the acceleration is constructed for each data point.

$$v = [u^T, \omega^T]^T = [a_x^f, a_y^f, a_z^f, \omega_x^f, \omega_y^f, \omega_z^f, d_x, d_y, d_z]^T$$

B. HMM training phase

Each strong displacement activity is represented by one HMM model trained by a series of data recorded when the human subject repeatedly performs the same activity in the training phase. Each parameter set of an HMM is trained by the data blocks with a specific label number indicating the type of activity recorded prior. The centroid of the K-means clustering is also trained for vector quantization and saved for the recognition phase.
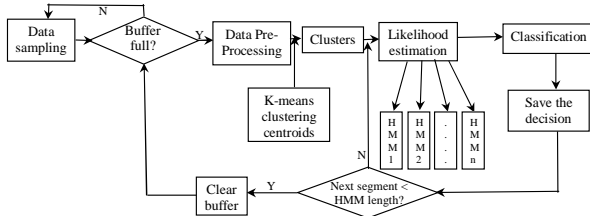

Fig. 4.   The flow chart of the HMM-based activity recognition algorithm.

C. HMM-based recognition phase

After the neural network module and sensor data fusion, the data is classified into zero displacement activities, transitional activities, and strong displacement activities. The data block of strong displacement activities is fed into the HMM-based activity recognition module. As shown in Figure 4, the data pre-processing is applied to the gesture data and the centroid trained in the training phase is used afterward to quantify the vectors into observable symbols. Then a sliding-window of 1 second length is moving along the symbol sequence and the likelihood under each set of HMM parameters is estimated. We choose the model which maximizes the likelihood over other HMMs to be the recognized type as the output decision of the sliding-window. Thus, this HMM-based recognition phase

gives a series of decisions on the activity based on the trained parameters of HMMs.

V. Experimental Results

In our experiments, the human subject wore two sensors: one on the right foot and the other on the waist as shown in Figure 1. Regular daily activities were performed: standing, sitting, walking level, walking upstairs, walking downstairs, running, sleeping, etc. We recorded five sets of data for the training purpose and five sets for the recognition testing purpose.

A. Evaluation of the NN-based coarse-grained classification

The first and the second layers of the neural network are trained through MATLAB Neural Network Toolbox. The number of neurons in each layer is determined by balancing the training iterations and the performance index of the network. The initial values of the weights and biases are randomly selected, which will lead the performance of the network approach a local minimal. Within 300 iterations and a goal of 0.05, different initial values achieve various performances. The performance is monitored in order to achieve good training results. If the performance curve does not meet the goal, the training has to be restarted. The neural network NN_w for the waist and the neural network NN_f for the foot are trained separately using the data from its corresponding sensor.

Figure 5 shows good and bad training results of the neural networks, respectively. Only when the performance curve goes below the goal, the network can achieve adequate accuracy and a few error points scattered on the edges of the blocks. If the training goal has not been met, there will be consecutive errors which cause errors in sensor fusion.

TABLE II

DECISION ACCURACY OBTAINED FROM THE TESTING DATA.

Activity Type	HMM decision Type			
	Walking	Walking downstairs	Walking upstairs	running
Walking	**0.9030**	0.0581	0.0360	0.0029
Walking downstairs	0.0478	**0.9250**	0.0270	0.0020
Walking upstairs	0.0759	0.0289	**0.8915**	0.0037
running	0.0901	0.0120	0.0278	**0.8701**
Accuracy	**0.9030**	**0.9250**	**0.8915**	**0.8701**

B. Evaluation of the HMM-based recognition algorithm

Based on the results of the neural network, the hidden Markov model block is switched on when there is a cyclic activity. A sliding-window moves along the segmented data with the length of 1 second and step length of 0.2 second. The output of the sliding-window is a sequence of classification decisions. Then, a voting function follows to produce a single decision for each 1 second of time period. The HMM-based recognition results on the testing data after the voting function are shown in Table II. The percentages of decision under each ground truth are listed in each row, where the values on the diagonal indicate the accuracy of each activity.

The final result is a sequence of decisions corresponding to the time. For example, Figure 6 shows the raw angular velocity (top), and the decision results after the voting function compared with the ground truth (bottom). In the top figure, the 3D angular velocity from the sensor indicates several cyclic

Performance is 0.0381045, Goal is 0.05

Performance is 0.0424468, Goal is 0.05

Performance

Goal

300 Iterations

Performance

Goal

300 Iterations

Neural network output for the foot sensor

Neural network output for the waist sensor

Ground truth

Labels: 1. Standing/sitting, 2. sleeping, 3, sit-to-sleep, sleep-to-sit
4.Walking level, 5.Walking downstairs, 6.Walking upstairs, 8. running

Ground truth

Labels: 1. Standing/sitting, 2. sleeping, 3, sit-to-sleep, sleep-to-sit
4.Walking, 6. stand-to-sit, 7. sit-to-stand, 8. running
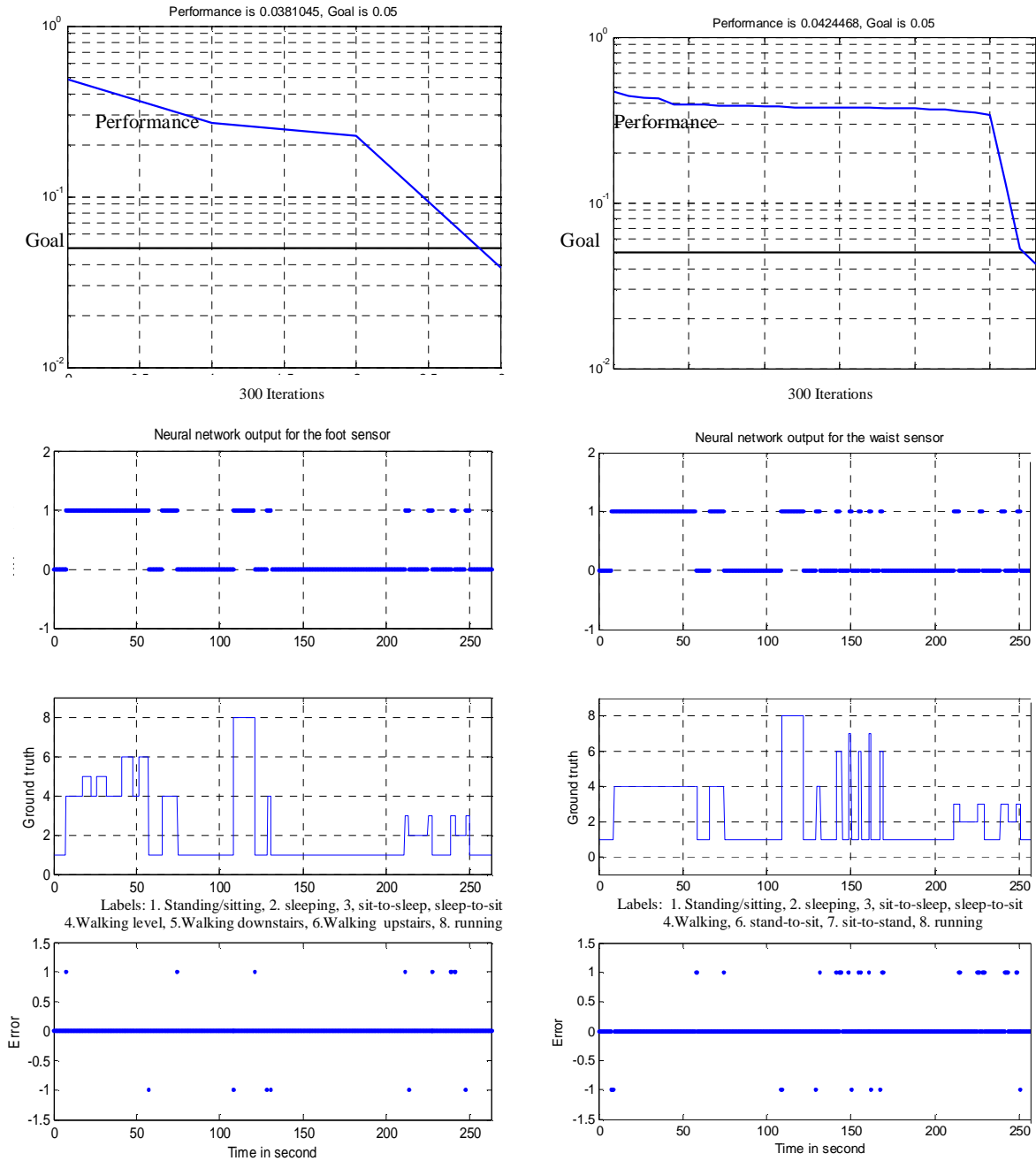
Error

Time in second

Error

Time in second

Fig. 5. Left: The performance goal of the foot sensor was met, accuracy =98.40%. Right: The performance goal of the foot sensor was met, accuracy =94.61%.

activities, transitional activities, and stationary activities. In the bottom figure there are several misclassifications in the circled areas. With the heuristic method of the segmentation refinement module in the sensor fusion function, a whole segmentation is produced rather than several short ones. The two circles on the bottom figure show that the neural networks and the sensor fusion give correct segmentation output and the errors are caused by the HMM-based recognition algorithm.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we introduced a robot-assisted living system for elderly people, patients, and the disabled. We proposed a multi-sensor fusion-based activity recognition algorithm which combines neural networks and hidden Markov models. The HMM-based recognition algorithm is applied only to strong displacement activities. Therefore, the calculation complexity has been reduced and the efficiency of the algorithm is enhanced by the fusion of the data from these two sensors. In the future, we will implement the online recognition algorithm
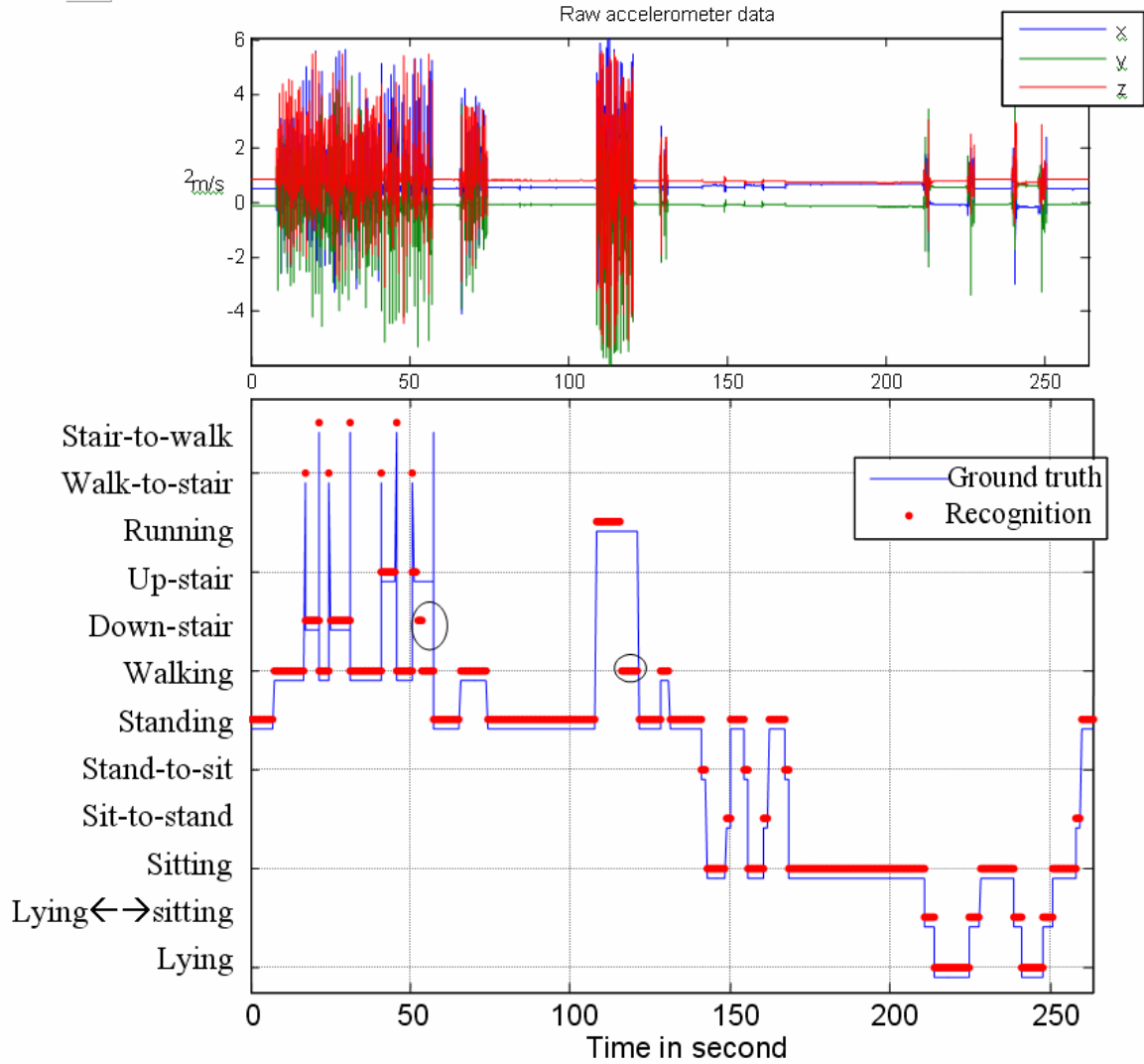
Fig. 6. The final results of the activity classification.

and utilize other sensor nodes to improve the accuracy and increase the variety of activities.

REFERENCES

[1] *Baby Boomers Aging Needs*. http://www.Babyboomercaretaker.com, Oct 2008.
[2] Z. Khalila and M. Merhia. Effects of aging on neurogenic vasodilator responses evoked by transcutaneous electrical nerve stimulation. *Journals of Gerontology Series*, pages B257–B263, 2000.
[3] W. C. Mann. *Smart Technology for Aging, Disability, and Independence*. John Wiley Sons, Inc., 2005.
[4] C. Zhu, W. Sun, and W. Sheng. Wearable sensors based human intention recognition in smart assisted living systems. In *IEEE International Conference on Information and Automation*, pages 954 – 959, 2008.
[5] C. Zhu, Q. Cheng, and W. Sheng. Human intention recognition in smart assisted living systems using a hierarchical hidden markov model. *IEEE International Conference on Automation Science and Engineering*, pages 253–258, 2008.
[6] G.Z. Yang and M. Yacoub. *Body Sensor Networks*. Springer, 2006.
[7] A. Ohta and N. Amano. Vision-based human behavior recognition by a mobile robot. *SICE. Annual Conference*, pages 3047–3051, 2007.
[8] T. B. Moeslunda, A. Hiltonb, and V. Kruger. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 104:90–126, 2006.
[9] K. Sagawa, T. Ishihara, A. Ina, and H. Inooka. Classification of human moving patterns using air pressure and acceleration. *Industrial Electronics Society, 1998. IECON '98. Proceedings of the 24th Annual Conference of the IEEE*, 2:1214 – 1219, 1998.
[10] J. Mantyjarvi, J. Himberg, and T. Seppanen. Recognizing human motion with multiple acceleration sensors. *Systems, Man, and Cybernetics, 2001 IEEE International Conference on*, 2:747 – 752, 2001.
[11] J. Lester, T. Choudhury, and G. Borriello. A practical approach to recognizing physical activities. *Joint SOC-EUSAI conference*, pages 1 – 16, 2006.
[12] O. Amft, H. Junker, P. Lukowicz, G. Troster, and C. Schuster. Sensing muscle activities with body-worn sensors. *Wearable and Implantable Body Sensor Networks International Workshop on*, page 4, 2006.
[13] T. M. Mitchell. Machine learning. *Boston WCB/McGraw-Hill*, 1997.
[14] K. Aminian, Ph. Robert, E. E. Buchser, B. Rutschmann, D. Hayoz, and M. Depairon. Physical activity monitoring based on accelerometry: validation and comparison with video observation. *Medical and Biological Engineering and Computing*, 3:304–308, 1999.
[15] T. Mitchell. Decision tree learning. *Machine Learning*, pages 52–78, 1997.
[16] D. Lowd and P. Domingos. Naive bayes models for probability

estimation. *Proceedings of the 22 nd International Conference on Machine Learning*, 2005.

[17] L. R. Rabiner. A tutorial on hidden markov models and selected application in speech recognition. In *Proc. IEEE*, volume 77, pages 267–296, 1989.

[18] R. W. DeVaul and S. Dunn. Real-time motion classification for wearable computing applications. *Technical report, MIT Media Laboratory*, 2001.

[19] J. Lester, T. Choudhury, N. Kern, G. Borriello, and B. Hannaford. A hybrid discriminative/generative approach for modeling human activities. In *In Proc. of the International Joint Conference on Artificial Intelligence IJCAI*, pages 766–772, 2005.

[20] Y. Freund. Boosting a weak learning algorithm by majority. *Proceedings of the Third Annual Workshop on Computational Learning Theory*, pages 202–216, 1990.

[21] M. T. Hagan, H. B. Demuth, and M. H. Beale. *Neural Network Design*. PWS Publishing Company, 1996.