Correspondence

Wearable Sensor-Based Hand Gesture and Daily Activity Recognition for Robot-Assisted Living

Chun Zhu and Weihua Sheng

***Abstract*—In this paper, we address natural human-robot interaction (HRI) in a smart assisted living (SAIL) system for the elderly and the disabled. Two common HRI problems are studied: hand gesture recognition and daily activity recognition. For hand gesture recognition, we implemented a neural network for gesture spotting and a hierarchical hidden Markov model for context-based recognition. For daily activity recognition, a multisensor fusion scheme is developed to process motion data collected from the foot and the waist of a human subject. Experiments using a prototype wearable sensor system show the effectiveness and accuracy of our algorithms.**

***Index Terms*—Assisted living, hidden Markov models (HMNs), human-robot interaction (HRI), wearable computing.**

I. Introduction

A. Motivation

The past decade has seen a steady growth of the elderly population. Compared with the rest of the population, more seniors live alone as sole occupants of a private dwelling than any other population group. Helping them to live a better life is very important and has great societal benefits. Many researchers are working on new technologies such as assistive robots to help elderly people. Haigh and Yanco [1] provided a survey on assistive robots used as caregivers. Recently, we have developed a *smart assisted living* (SAIL) system [2], [3] to provide support to elderly people in their residence. As illustrated in Fig. 1, the SAIL system consists of a body sensor network (BSN), a companion robot, a smartphone, and a remote health provider. The BSN collects motion data and vital signs from a human subject and sends them wirelessly (for example, through Zigbee) to the companion robot, which infers the human intentions and health conditions from these data and responds accordingly. The smartphone serves as a gateway to access the expertise of remote health-care providers, if needed.

Natural human-robot interaction (HRI) is a very important issue in the design of assistive robotics, particularly for elderly people, who usually suffer from problems with speech or have difficulty in learning new computer skills. It is desirable to make the robot able to not only understand explicit human intentions from gestures but also recognize implicit intentions inferred from daily activities. Such a robot capability is called *considerate intelligence* [2]. In this paper, we focus on solving two problems central to natural HRI: hand gesture recognition and daily activity recognition. Compared with the existing work, we made two main contributions. First, we developed a lightweight and resource-aware hand gesture recognition algorithm that considers the context information or the sequential constraints between different gestures. Additionally, the algorithm can automatically spot the start and end points of a gesture. Second, we developed a multisensor fusion scheme for accurate daily activity recognition. Motion data from the foot and waist sensors are fused to recognize 13 daily activities such as standing, sitting, and lying.

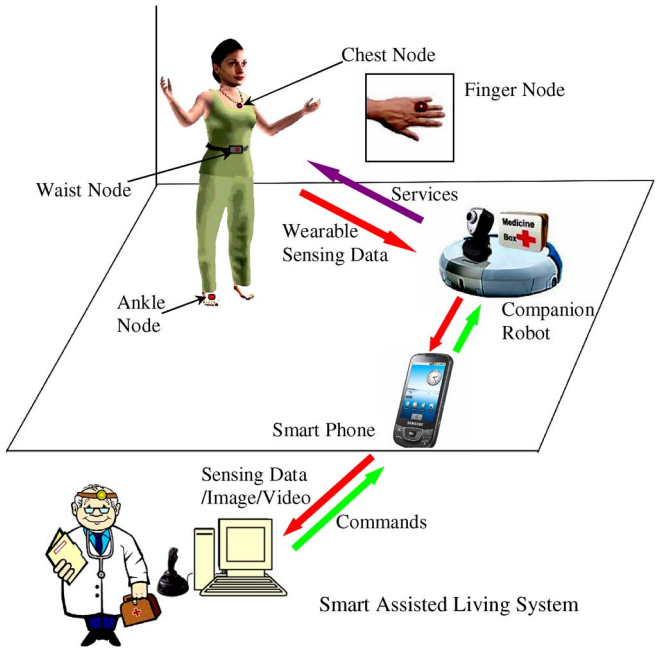

Fig. 1. Overview of the SAIL system.

This paper is organized as follows. The rest of this section introduces some related work in hand gesture and daily activity recognition. Section II develops the algorithm for hand gesture recognition. Section III describes the algorithm for daily activity recognition. The experimental tests and results are presented in Section IV. Conclusions are given in Section V.

B. Related Work

Researchers have made significant progress in the area of HRI in recent years. A comprehensive survey of this area is provided by Yanco and Drury [4]. Hand gesture recognition and human daily activity recognition are essential to natural HRI. Traditional gesture and daily activity recognition is based on visual input [4]. A typical approach has two steps: 1) feature extraction using color detection, edge detection, background-removing techniques, etc., and 2) pattern recognition using machine-learning algorithms. More works in this area can be found in [5] and [6]. Recently, due to the advancement in microelectromechanical system and very large scale integration technologies, wearable sensor-based gesture and daily activity recognition has been gaining attention. Wearable sensors have less dependence on their surroundings, and gesture recognition systems based on wearable sensors require less data compared to vision-based systems.

An important problem in gesture recognition is to segment gesture from nongesture movements, which is called the *gesture spotting problem* [7]. There are two main solutions to this problem: rule-based [8] and hidden Markov model (HMM)-based methods [9]. Rule-based methods are easy to implement, but they have special hand movement requirement for the human subject, which is not convenient in practical

Manuscript received January 8, 2009; revised August 13, 2009 and May 18, 2010; accepted June 4, 2010. This paper was recommended by Associate Editor Q. Ji.

The authors are with the School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK 74078 USA (e-mail: chunz@okstate.edu; weihua.sheng@okstate.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMCA.2010.2093883

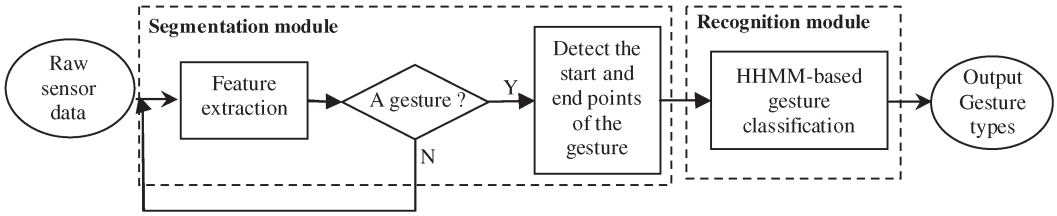

Fig. 2. Flowchart of the hand gesture recognition algorithm.

use. HMM-based methods calculate the likelihood of gestures on a sliding window. However, the computational cost is high due to the use of HMMs. In this paper, we adopted a new method in our hand gesture recognition to avoid special requirement for the human subject and heavy computational cost.

Over the years, many solutions have been developed for daily activity recognition, including heuristic analysis methods [10], discriminative methods [11], generative methods [12], and some combinations of these methods [13]. Heuristic analysis methods require intuitive analysis of the raw sensor data or the features from data. Discriminative methods and generative methods are machine-learning algorithms. Parameters are trained using data from different individuals, but the computational cost is usually high. In this paper, we use a combination of discriminative and generative methods to achieve better performance for daily activity recognition.

II. Hand Gesture Recognition

In our SAIL system, different hand movement patterns are used to command the companion robot, much like the way we command a dog. Five basic hand gestures are assigned to five commands, which mean "come," "go fetching," "go away," "sit down," and "stand up," respectively. Here, we discuss our algorithm for hand gesture recognition, which combines neural network (NN)-based gesture spotting and hierarchical HMM (HHMM)-based gesture classification.

Since most embedded computing systems have limited battery power and computational power, it is important to design recognition algorithms that are resource-aware and lightweight. As shown in Fig. 2, our recognition algorithm consists of two modules: 1) a segmentation module that uses an NN to realize gesture spotting and 2) a recognition module that uses an HHMM to classify gestures. Since the HHMM is a probabilistic model with high computational cost, the NN-based segmentation module is used as a switch to control the data flow in order to save computation time and increase efficiency.

A. Gesture Spotting Using an NN

A three-layer feedforward NN is implemented to distinguish gestures from daily nongesture movements. We find that simply using a single threshold on the sensor data cannot distinguish gestures and nongesture movements accurately. On the contrary, the NN is a combination of multiple thresholds for different features. Through the training of the NN, the weights and biases can be optimized for classification. Furthermore, the NN is a machine-learning algorithm, which can obtain hidden information from the training data and make an accurate combination of features to perform the classification for gestures and nongesture movements.

The input to the NN is the feature vector extracted from the raw sensor data. We use the means and variance as the features. The output is binary (1 or 0), which stands for gestures or nongesture movements, respectively.

Supervised learning is used to train the NN. In the training mode, the human subject labels the ground truth when he/she is performing gestures and nongesture movements. The back-propagation method is implemented to train the weights and biases of the neurons in the first and second layers.

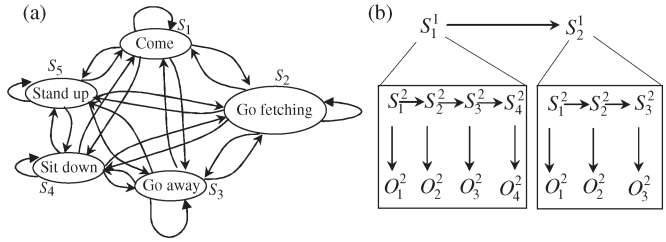

Fig. 3. (a) Transition of the upper level HMM that considers the context information. (b) Architecture of an HHMM.

In our current implementation, we assume that nongesture movements are slow because when people read, write, walk, and eat, their hands do not exhibit intensive motions. For unexpected movements and rapid nongesture movements, a threshold-based HMM likelihood discriminant [14] can be used to distinguish whether it is a gesture.

B. HHMM-Based Recognition Algorithm

People usually demonstrate specific patterns when they interact with their pets, and such patterns reflect the sequential constraints on the gestures, which can be used to improve gesture recognition accuracy. In this paper, the HHMM technique is implemented for this purpose. The HHMM is a statistical model derived from the HMM and can be used to represent sequential constraints. The HMM has been widely used in speech recognition, handwriting recognition, and other pattern recognition applications. The HMM can be applied to represent the statistical behavior of an observable symbol sequence in terms of a network of states. We recognize gestures through two steps: 1) using the HMMs at the lower level to recognize individual hand gestures and 2) modeling the constraints on the gestures with the upper level HMM and estimating the most likely state sequence in the upper level HMM to correct classification errors made in the lower level HMM.

1) HMM-Based Individual Hand Gesture Recognition: We preprocessed raw sensor data (3-D acceleration and 3-D angular velocity) to extract the features (means and variance) for gesture classification in the lower level HMM, which has two phases: 1) the training phase and 2) the recognition phase.

In the training phase, there are four steps:

Step 1: Find the stroke duration using fast Fourier transform (FFT).
Step 2: Quantify the vectors into observation symbols using the K-means clustering.
Step 3: Set up the initial HMM parameters.
Step 4: Iterate for the EM method [15].

In the recognition phase, the likelihood of the sequential data was estimated for each set of HMM parameters. We chose the model that obtains the maximum likelihood to be the recognized type.

2) Context-Based Hand Gesture Recognition: After individual hand gestures are recognized, the upper level of the HHMM is used to consider the sequential constraints between the gestures or the *context.* Fig. 3(a) shows the transition of the upper level HMM. It undergoes a

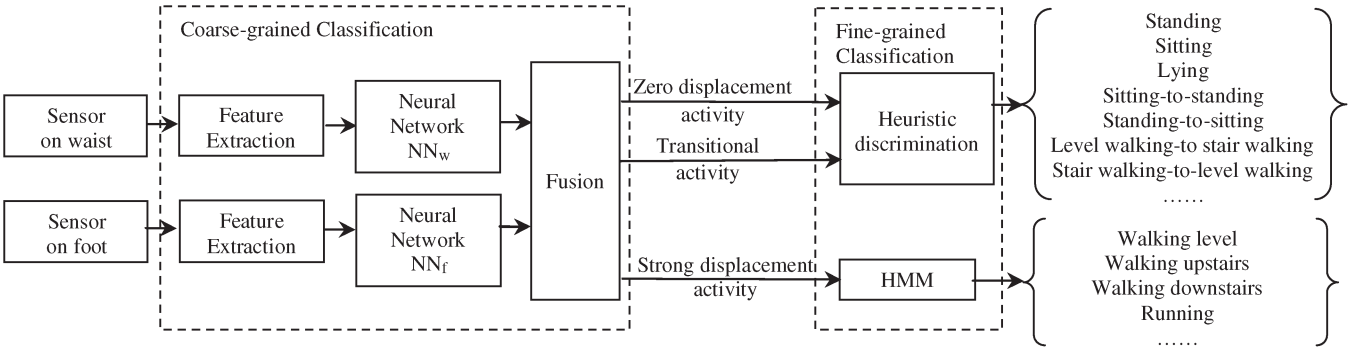

Fig. 4. Overview of the daily activity recognition algorithm.

change of state according to a set of probabilities associated with each state. For example, the same command is less likely to be sent twice back to back, and when the previous command is "go away," the next one has a small probability of being "go fetching."

The HHMM is a generalization of the segment model where each segment has subsegments. Fig. 3(b) illustrates the basic idea of an HHMM. A time series is hierarchically divided into segments, where S_i^1 represents the state at the upper level HMM, and S_i^2 represents the state at the lower level HMM.

We denote the time slice associated with the state change as $k = 1, 2, \ldots$, and the kth actual state as s_k. For a given observation sequence with a length of T, the Viterbi algorithm [16] is used at the upper level HMM to find the single best state sequence $Q = \{s_1, s_2, \ldots, s_T\}$, which represents the most likely underlying gesture sequence for the given observation sequence $O = \{O_1, O_2, \ldots, O_T\}$. This way, some errors in the lower level HMM can be corrected by the upper level HMM.

III. Daily Activity Recognition

Here, we discuss the human daily activity recognition algorithm through multisensor fusion. Two inertial sensors are attached to one foot and the waist of the subject, respectively. As shown in Fig. 4, the raw sensor data (acceleration and angular velocity) are processed to obtain the features (mean, variance, and covariance), which are fed into two NNs NN_f and NN_w for foot and waist, respectively. The fusion of the results from the two NNs generates coarse-grained classification for three types of human activities: zero-displacement activities, transitional activities, and strong displacement activities. Then, a heuristic discrimination module is used for fine-grained classification of zero-displacement activities and transitional activities, and an HMM-based recognition algorithm is used for fine-grained classification of strong displacement activities. 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. This mechanism can save the computation time and enhance the efficiency of the recognition algorithm.

A. Coarse-Grained Classification

We categorize the daily activities into the following three types: 1) zero-displacement activities $A_Z = \{standing,\ sitting,\ lying\}$; 2) transitional activities $A_T = \{sitting\text{-}to\text{-}standing,\ standing\text{-}to\text{-}sitting,\ level\ walking\text{-}to\text{-}stair\ walking,\ stair\ walking\text{-}to\text{-}level\ walking,\ lying\text{-}to\text{-}sitting, sitting\text{-}to\text{-}lying\}$; and 3) strong displacement activities $A_S = \{walking\ level,\ walking\ upstairs,\ walking\ downstairs,\ running\}$.

The NNs NN_f and NN_w classify the foot activity A_f and waist activity A_w into three types: 1) *stationary*, 2) *transitional*, and 3) *cyclic*. The outputs of the NNs are fed into the fusion module, which integrates the individual types of foot and waist activities and categorizes human activities according to the following rules: 1) zero-displacement activities: $A \in A_Z$ if and only if $A_w = stationary$; 2) transitional: $A \in A_T$ if and only if ($A_f = transitional$ and $A_w = transitional$) or ($A_f = stationary$ and $A_w = transitional$); and 3) strong displacement activities: $A \in A_S$ if and only if $A_f = cyclic$ and $A_w = cyclic$. All other combinations of foot and waist activities are considered as rare activities, and they are not included in this paper.

B. Fine-Grained Classification

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 heuristic discrimination module is applied to consider the previous stationary activity and decide the type of the current transitional activity. For example, when the detected previous activity is *sitting*, after a transitional activity, the following activity is stationary. Then, we use a discriminative model to test whether the direction of the torso is vertical or horizontal: If it is vertical, then the current activity is *standing*, and the previous transitional activity is *sitting-to-standing*; otherwise, the current activity is *lying*, and the previous transitional activity is *sitting-to-lying*.

An HMM-based recognition algorithm is applied to further determine the types of the strong displacement activities, which recognizes the patterns of the continuous time series of data. The detailed algorithm is similar to the upper level HMM used in hand gesture recognition, which will not be repeated here.

IV. Experimental Results

A. Hand Gesture Recognition

1) Experiment Setup and Process: For hand gesture recognition, we use an inertial sensor *nIMU* from MEMSense, LLC [17], which provides 3-D acceleration, angular velocity, magnetic data, and temperature data at a sampling rate of 150 Hz. The prototype of the wearable sensor system for hand gesture recognition is shown in Fig. 5(a). The *uIMU* sensor is connected to a personal digital assistant (PDA) through an RS422/RS232 serial converter, and the PDA sends the data to a desktop computer through Wi-Fi. In the experiments, we define the following five gestures as shown in Fig. 6:

- waving hand backward for "come here";
- waving left and right for "go away";
- pointing forward for "go fetching";
- turning clockwise for "sit down";
- turning counterclockwise for "stand up."

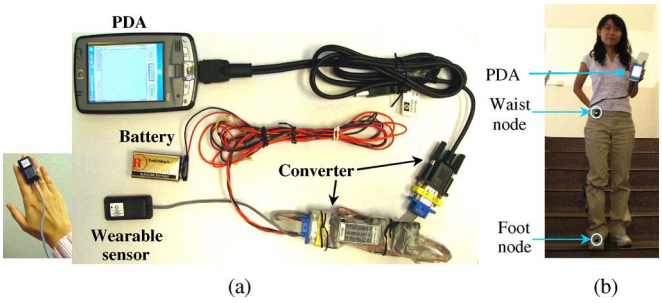

Fig. 5. (a) Prototype of the wearable sensor system for hand gesture recognition. (b) Experiment setup for daily activity recognition.

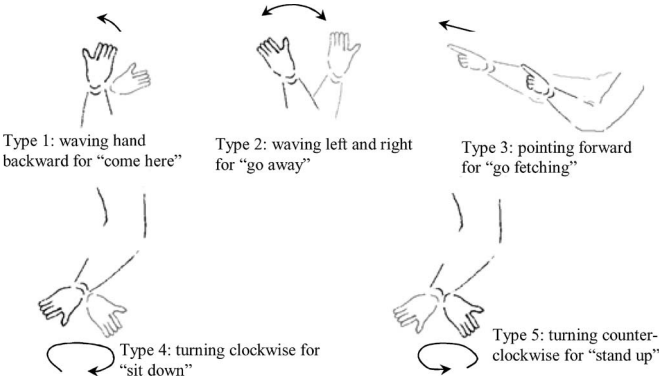

Fig. 6. Hand gestures for the five commands.

We recorded five sets of data for the training purpose and five sets for the testing purpose. In the experiments, we carried out the following three steps.

Step 1: Repeatedly perform gesture type 1 for 15 times and take a 5-s break. Continue performing the rest types following the same pattern until type 5 is done. Label each gesture and record the data on a file.

Step 2: Perform a sequence of 20 gestures with a break of about 3 s between gestures. The gestures mimic a real-world scenario of interacting with a robot.

Step 3: Process the training data and test data. First, train the NN to distinguish gestures from daily nongesture movements. Second, use the training data to train the lower level HMMs. Third, use the trained HMMs to recognize individual gestures in the test data. The output of each test is a sequence of recognized gestures. Finally, the Viterbi algorithm is used to produce the most likely underlying gesture sequence based on the given upper level HMM.

2) Gesture Recognition Result: The parameters of the upper level HMM are obtained by observing the human subject interacting with the robot for a sustained period of time (for example, one day). The transition matrix can be different from person to person. The observation symbol probability distribution matrix is equivalent to the accuracy matrix of each individual gesture in the lower level HMMs, which can be obtained from the individual gesture recognition. The initial state distribution is set to be a uniform distribution to reflect the fact that no preference will be given to a specific command.

Fig. 7 shows the recognition results of the testing data. In subfigure (a), the 3-D angular velocity from the sensor is shown. In subfigure (b), the NN helps to spot the gestures. In subfigure (c), when the lower level HMMs are applied, there are some errors at points a, b, c, d, e, and f. In subfigure (d), after considering the context information, the errors at b, c, and f are corrected by the Bayesian filtering in the upper level of HHMM. The performance of recognition is evaluated by comparing the result with the ground truth. The decisions and the classification accuracy of the HMM-based and HHMM-based recognition are listed in Tables I and II, respectively. Each row shows the number of decisions in each category given the ground truth of the gesture. It is obvious that the performance of HHMM is much better than that of individual HMMs only. For the video clips of the experiments, please see the link in [18].

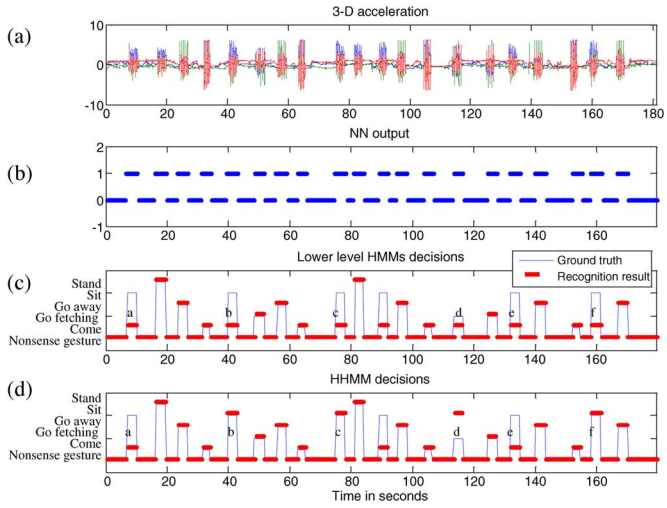

Fig. 7. Results of the NN and HMMs. (a) Raw angular velocity. (b) Output of the NN. (c) Individual HMM recognition results compared with the ground truth. (d) HHMM recognition results compared with the ground truth.

TABLE I
ACCURACY OF HMM-BASED RECOGNITION

Ground Truth	Decision Type 1	2	3	4	5	Test Accuracy
1	95	2	2	2	1	0.9314
2	3	82	15	1	0	0.8119
3	0	9	99	1	1	0.9000
4	41	4	0	54	1	0.5400
5	6	0	1	0	96	0.9320

TABLE II
ACCURACY OF HHMM-BASED RECOGNITION

Ground Truth	Decision Type 1	2	3	4	5	Test Accuracy
1	99	1	0	2	0	0.9706
2	0	86	15	0	0	0.8515
3	0	7	103	0	0	0.9364
4	10	2	2	86	0	0.8600
5	2	0	0	0	101	0.9806

B. Daily Activity Recognition

1) Experiment Setup and Process: For daily activity recognition, we use two inertial sensors. The experiment setup is shown in Fig. 5(b). Both inertial sensors are connected to a PDA through RS422/RS232 serial converters. The PDA sends data to a desktop computer through Wi-Fi. In our experiments, regular daily activities were performed: *standing*, *sitting*, *walking level*, *walking upstairs*, *walking downstairs*, *running*, *lying*, etc. We recorded five sets of data for the training purpose and five sets for the testing purpose. The NN NN_w for the waist and NN_f for the foot are trained separately with the data collected by the corresponding sensors. When the training performance is satisfactory, the NN can achieve adequate accuracy and only a few errors on the start and end points of each activity block.

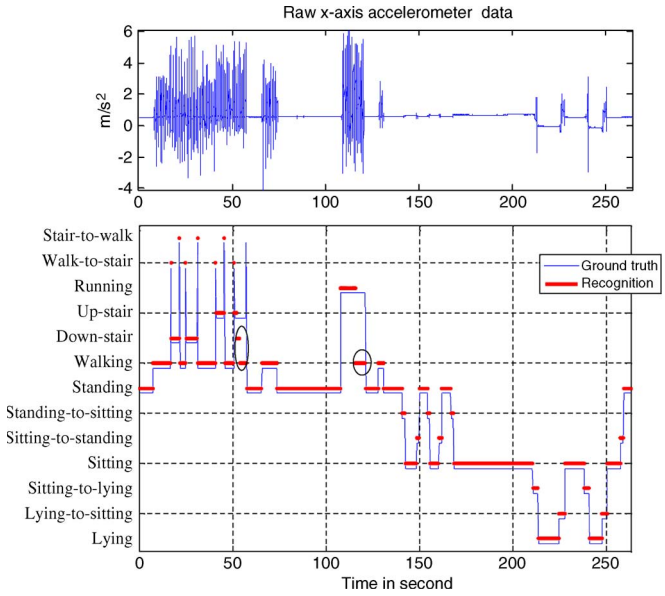

Fig. 8. Final results of the daily activity recognition.

TABLE III
CLASSIFICATION ACCURACY OBTAINED FROM THE TESTING DATA

Activity Type	HMM decision type				Test Accuracy
	Walking	Walking downstairs	Walking upstairs	Running	
Walking	623	40	36	2	0.8887
Walking downstairs	24	463	13	5	0.9168
Walking upstairs	41	16	481	8	0.8810
Running	23	3	7	218	0.8685

2) Activity Recognition Result: Based on the results of the coarse-grained classification, the heuristic discrimination module or the HMM-based recognition module will be applied for fine-grained classification. Our tests show that the accuracy of the heuristic discrimination module is about 98.3%. The HMM module is switched on when there is a strong displacement activity. A sliding window moves along the segmented data with a length of 1 s and a step length of 0.2 s. The output is a sequence of classification decisions. Then, a simple majority voting function follows to produce a single decision for each activity segment.

Fig. 8 shows the final results of the daily activity classification. The upper subfigure shows the x-axis acceleration from the sensor. In the lower subfigure, there are several classification errors indicated by the circles. The two circles on the lower subfigure show that the errors are caused by the HMM-based recognition algorithm for the strong displacement activities. The HMM-based recognition results on the testing data are shown in Table III. The integers are the total number of decisions made for each given activity type.

V. CONCLUSION

In this paper, we have introduced a SAIL system for the elderly and the disabled. To realize natural HRI in such a SAIL system, we have proposed 1) an NN-based gesture spotting and an HHMM-based hand gesture recognition algorithm, and 2) a multisensor fusion-based daily activity recognition algorithm. For hand gesture recognition, an HHMM has been used to model the sequential constraints on the gestures, which increases the recognition accuracy. For daily activity recognition, the multisensor fusion scheme can increase the types of daily activities to be recognized. Fusion of the outputs of the NNs for the foot sensor and the waist sensor produces coarse-grained classification. Then, in the fine-grained classification, the HMM module has only been applied to strong displacement activities, whereas the heuristic discrimination module has been applied to both zero-displacement activities and transitional activities. Our experimental results have verified the accuracy of the algorithms. In the future, we will implement the recognition algorithms on a real robot to enable real-time HRI.

REFERENCES

[1] K. Z. Haigh and H. Yanco, "Automation as caregiver: A survey of issues and technologies," in *Proc. AAAI Workshop "Automation as Caregiver"*, 2002, pp. 39–53, AAAI Tech. Rep. WS-02-02.
[2] C. Zhu, W. Sun, and W. Sheng, "Wearable sensors based human intention recognition in smart assisted living systems," in *Proc. IEEE Int. Conf. Inf. Autom.*, 2008, pp. 954–959.
[3] C. Zhu, Q. Cheng, and W. Sheng, "Human intention recognition in smart assisted living systems using a hierarchical hidden Markov model," in *Proc. IEEE Int. Conf. Autom. Sci. Eng.*, 2008, pp. 253–258.
[4] H. A. Yanco and J. L. Drury, "Classifying human-robot interaction: An updated taxonomy," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, 2004, pp. 2841–2846.
[5] K. Abe, H. Saito, and S. Ozawa, "Virtual 3-D interface system via hand motion recognition from two cameras," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 32, no. 4, pp. 536–540, Jul. 2002.
[6] M. De Marsico, M. Nappi, and D. Riccio, "FARO: Face recognition against occlusions and expression variations," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 40, no. 1, pp. 121–132, Jan. 2010.
[7] R. Oka, "Spotting method for classification of real world data," *Comput. J.*, vol. 41, no. 8, pp. 559–565, 1998.
[8] A. Ramamoorthy, N. Vaswani, S. Chaudhury, and S. Banerjee, "Recognition of dynamic hand gestures," *Pattern Recognit.*, vol. 36, no. 9, pp. 2069–2081, Sep. 2003.
[9] L. R. Rabiner, "A tutorial on hidden Markov models and selected application in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.
[10] B. Najafi, K. Aminian, A. Paraschiv-Ionescu, F. Loew, C. J. Bula, and P. Robert, "Ambulatory system for human motion analysis using a kinematic sensor: Monitoring of daily physical activity in the elderly," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 6, pp. 711–723, Jun. 2003.
[11] J. P. Wachs, H. Stern, and Y. Edan, "Cluster labeling and parameter estimation for the automated setup of a hand-gesture recognition system," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 35, no. 6, pp. 932–944, Nov. 2005.
[12] J. Yang, Y. Xu, and C. S. Chen, "Human action learning via hidden Markov model," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 27, no. 1, pp. 34–44, Jan. 1997.
[13] J. Lester, T. Choudhury, N. Kern, G. Borriello, and B. Hannaford, "A hybrid discriminative/generative approach for modeling human activities," in *Proc. IJCAI*, 2005, pp. 766–772.
[14] H. K. Lee and J. H. Kim, "An HMM-based threshold model approach for gesture recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 10, pp. 961–973, Oct. 1999.
[15] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc.*, vol. 39, no. 1, pp. 1–38, 1977.
[16] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimal decoding algorithm," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 2, pp. 260–269, Apr. 1967.
[17] LLC MEMSense, 2009. [Online]. Available: http://www.memsense.com/
[18] C. Zhu, Hand Gesture Recognition for Human Robot Interaction (HRI), 2009. [Online]. Available: http://www.youtube.com/watch?v=ypq14K2HMMQ