Proceedings of the 2010 IEEE
International Conference on Information and Automation
June 20 - 23, Harbin, China

Realtime Human Daily Activity Recognition Through Fusion of Motion and Location Data

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

***Abstract*—As robot assisted living is gaining more attentions for elderly care recently, automated human daily activity recognition becomes more important in human-robot interaction. 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 collect motion data, while an optical motion capture system is used to record 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 using the neural networks and a hidden Markov model. 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.**

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

I. Introduction

A. Motivation

In recent years, elderly care is gaining more and more attention due to the growth of elderly population. In order to help seniors live a better life, there is an urgent requirement for robot assisted living systems to help elderly people. We are developing a smart assisted living (SAIL) system [1], [2], which can help elderly people in their daily life. Automated recognition of human daily activities is very important for human-robot interaction (HRI) [3] in assisted living systems. In addition, human daily activity recognition can help people studying behavior related diseases and detecting abnormal behaviors.

Visual data is widely used in activity recognition because vision-based systems can observe full human body movements but have the data association problem for multiple human subjects. It is very complicated to recognize human activities by processing image data. In addition, visual data is easy to be influenced by environments. Recently, researchers have become interested in using wearable sensors, wearable sensors can track motions with less data compared to vision-based systems. Since too many wearable sensors are uncomfortable and obtrusive to users, it is crucial to build a minimum wearable sensor system to recognize human daily activities.

In this paper, we proposed an approach that combines motion data and vision-based location information to recognize human daily activities. This approach has the following advantages: first, a single inertial sensor worn by the user for motion data collection can reduce the 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 by considering the sequential constraints in daily activities.

This paper is organized as follows. The rest of Section I introduces the related work in this area. Section II describes the hardware platform for the proposed human daily activity recognition system. Section III explains activity recognition from motion data only. Section IV presents the fusion of motion data and location information in a Bayesian framework. The experimental results are provided in Section V. Conclusions and future work are given in Section VI.

B. Related Work

Recently, human activity recognition is a popular area and many significant progresses have been made. Here, we will give a brief overview of both vision-based and wearable sensor-based human daily activity recognition methods.

Visual-based methods are the traditional approach to human daily activity recognition. Visual information processing for activity recognition deals the spatio-temporal interaction among the trajectories of different human body parts [4]. It is very important to detect and track the body parts in order to recognize the type of activities. A typical approach for vision-based recognition has two steps: feature extraction and pattern recognition. In the feature extraction step [5], activities are analyzed in terms of the trajectories of moving bounding boxes, and features are extracted from each image frame. In the pattern recognition step [6], activities are analyzed using context information of the body parts, which

978-1-4244-5704-5/10/$26.00 ©2010 IEEE

is represented by the extracted features. However, vision-based activity recognition incurs a significant amount of computational cost, and vision data are usually compromised by the environments, such as poor lighting conditions and occlusion.

Since wearable sensor-based systems have less data to process and it is easy to use, wearable sensor-based activity recognition has been gaining attention. Inertial sensors are usually used to capture human daily motion data. Many applications for human activity recognition using inertial sensors can be found in [7], [8]. For example, Aminian *et al.* [8] used two inertial sensors on the chest and on the rear of the thigh and sampled the acceleration of the chest and the thigh. Activities such as sitting, standing, lying, and dynamic (walking) activities can be detected by features from different directions of the sensors. However, they cannot further distinguish the detailed types of the dynamic activities. Sensors with other modalities can be used to provide complementary information to motion data and detect variety of activities, such as air pressure sensor, microphones, and temperature sensors. For example, Sagawa *et al.* [9] discussed a method to classify human moving behaviors using one acceleration sensor and one air pressure sensor attached to the waist. However, there are some shortcomings of wearable sensor-based activity recognition. Wearable sensor systems are obtrusive and inconvenient to the human subject especially when there are many wearable sensors. On the other hand, a single sensor is usually not sufficient to distinguish the basic daily activities due to the inherited ambiguity. In a single-sensor system [7], transition activities are used to distinguish different stationary activities, such as transitions between sitting and standing. Because this method has no error correction function, a mis-detection of a transition activity will cause accumulated error in the following detection. In order to enhance the accuracy of activity recognition, location information can be used to exploit activity-location correlation. Liao *et al.* [10], [11] used data from a wearable GPS location sensor to identify a user's significant places, and learn to discriminate between the activities performed at these locations. They also use a sensor board including a 3-axis accelerometer, two microphones for recording speech and ambient sound, phototransistors for measuring light conditions, and temperature and barometric pressure sensors. However, they do not aim to track people in the indoor environment. In addition, most researchers solved offline activity recognition, while only a few [12] focused on real-time activity recognition in previous work, which is important for elderly care.

II. Hardware Platform

Our proposed hardware system for human daily activity recognition is shown in Figure 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 decreases 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.

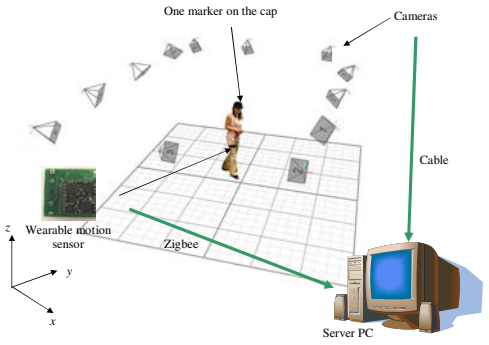

Fig. 1. The overview of the hardware platform for human daily activity recognition.

A. Hardware Setup for Motion Data Collection

Since the position to attach the sensor is very important to activity recognition[13], we collected data using the sensor on different parts of the human body and found that the thigh is the best location for activity recognition using the minimum sensor setup. As shown in Figure 2(a), a new motion sensor developed from a commercial product VN-100 [14] is attached to the right thigh of the human subject to collect motion data in this paper. Figure 2(b) and (c) shows the prototype of the wireless inertial sensor module. The VN-100 module can sense the 3D orientation, 3D acceleration, and 3D angular velocity and transfer the data to a desktop computer through the RF module XBee Module [15].

The wearable motion sensor samples the 3D acceleration and 3D angular velocity 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. Therefore we only collect the 3D acceleration as the raw data.

B. Hardware Setup for Location Information Collection

The OptiTrack motion capture system from NaturalPoint, Inc. [16] is marker-based and consists of twelve cameras. The tracking software runs on the server PC 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

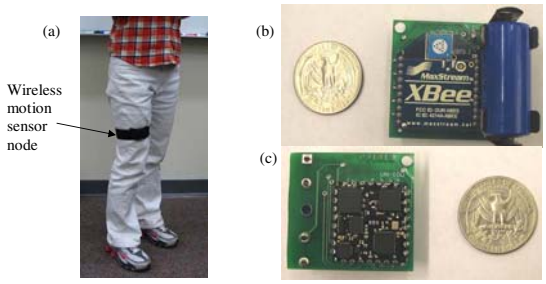

Fig. 2. The wireless inertial sensor module: (a)The wireless sensor module is worn on the thigh of the human subject; (b) XBee wireless module and the battery (front side); (c) VN-100 sensor module (back side).

100 fps. We down sample 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. 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.

III. Activity Recognition Using A Single Motion Sensor

We first develop a single wearable sensor-based activity recognition algorithm without considering the location information. Eight daily activities are to be detected: sitting, standing, lying, walking, sit-to-stand, stand-to-sit, lie-to-sit, and sit-to-lie. The activities are categorized into two kinds: stationary and motional activities. Figure 3 shows the classification of the eight activities into stationary and motional activities, and other activities. The number to the right of the activity is the activity ID.

There are two steps in the recognition algorithm from wearable sensor only: (1) coarse-grained classification and (2) fine-grained classification. The coarse-grained classification step uses the outputs of two neural networks and generates a basic classification result. The fine-grained classification step can realize real-time activity recognition with the sequential constraints modeled by an HMM using a modified short-time Viterbi algorithm [17] and generate the detailed activity types.

The following steps are used to recognize activities with motion data only. The detailed method can be found in [18].

1) A one-second window buffer is used to segment the motion data and extract features from the raw sensor data.
2) Two neural networks NN_1 and NN_2 are applied on different features of the sensor data to detect the state of the thigh: horizontal or vertical, and stationary or movement. Each neural network is a three-layer feed-forward network and is trained by labeled training data.

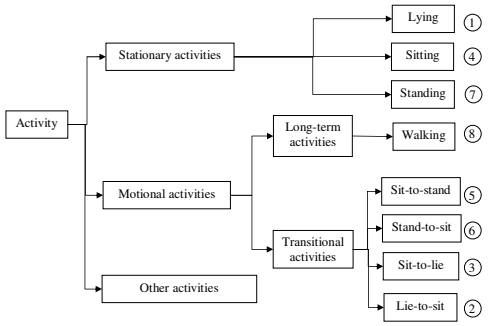

Fig. 3. The taxonomy of daily activities.

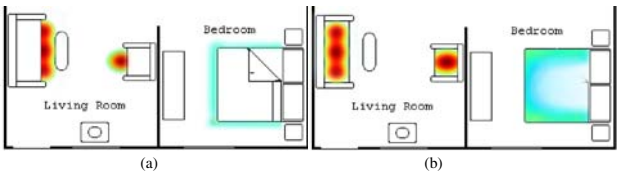

Fig. 4. The probability distribution of activities in the map: (a)"sitting" (b)"sit-to-stand".

3) A fusion function integrates the outputs of neural networks and produces the result of the coarse-grained classification.
4) A first order HMM is used to model sequential constraints in daily activities and a modified short-time Viterbi algorithm is applied to distinguish detailed activities by realtime. It is a decoding problem, which map the coarse-grained classifications to the fine-grained classifications. Since the standard Viterbi algorithm considers the whole observation sequence to find the best state sequence, which does not fit for real-time implementation, We proposed the modified short-time Viterbi algorithm, which can realize realtime daily activity decoding.

IV. 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 a floor plan of an apartment, we can infer the probability distribution for each specific activity on the 2D map. For example, Figure 4(a) shows the probability distribution of "sitting" and Figure 4(b) 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 Figure 5. Let $\hat{S}_i$ be the i^{th} estimated activity from the fine-grained classification step

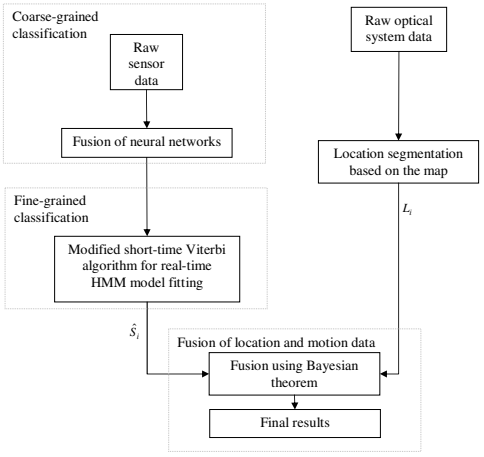

Fig. 5. The flow chart of the daily activity recognition algorithm.

and L_i be the corresponding location from the OptiTrack system. Bayes' theorem is used to fuse the motion data and the location information to obtain the final results. In reality, the coordinates of the human subject is a continuous function of time. 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 obtained this probability distribution function. First, it can be obtained based on a given floor plan in which locations and activities are correlated using human's knowledge, such as the situations in Figure 4(a) and (2). Second, it can be trained by observing the living pattern of a specific human subject for a sustained period of time, which is 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:

$$p(S_i|\hat{S}_i, L_i) \propto p(\hat{S}_i|S_i, L_i)p(S_i|L_i) \tag{1}$$

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:

$$p(\hat{S}_i|S_i, L_i) = p(\hat{S}_i|S_i) \tag{2}$$

$$p(S_i|\hat{S}_i, L_i) \propto p(\hat{S}_i|S_i)p(S_i|L_i) \tag{3}$$

where $p(\hat{S}_i|S_i)$ is the probability of observation distribution for each activity. $p(\hat{S}_i|S_i)$ represents the probability of recognition 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:

$$\hat{S}' = \arg\max_{S_i}(p(S_i|\hat{S}_i, L_i)) \tag{4}$$

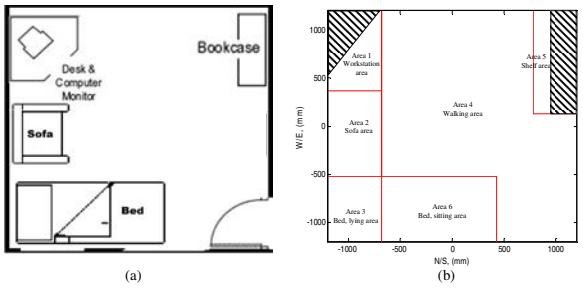

Fig. 6. (a)The layout of the mock apartment; (b)The segmentation of the room.

V. Experimental Results

A. Environment Setup

We performed the experiments in a mock apartment, which is in a lab environment with a dimension of 13.5 $\times$ 15.8 square feet as shown in Figure 6(a). The OptiTrack motion capture system is installed on the wall. To simplify the activity-location correlation, the given map of the mock apartment is segmented into different areas with corresponding probabilities of activity. The coordinate of the human subject given by the OptiTrack system is mapped into K semantic areas. The activity distribution given the area E can be represented by the conditional probability distribution function $p(S|E)$. 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 Figure 6(b), the room is segmented into 6 semantic areas: workstation area, sofa area, bed lying area, bed sitting area, 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.

As shown in Figure 1, the human subject wore the sensor on the right thigh and a cap with markers so that the head location can be tracked by the OptiTrack system. She moved slowly to mimic an elderly person's movement. The regular daily activities were performed: standing, sitting, sleeping, and transitional activities. We collected 5 sets of training data and 15 sets of testing data. Each testing data set had a duration of about 6 minutes. We recorded video in the meanwhile as the ground truth to evaluate the recognition results.

B. Evaluation of the Fusion of Location and Motion Data

For each second, an output decision value is generated in the experiment. A screen capture software is used to record the figures on the server PC, which shows the output of the recognition results. The captured results can be compared with the labeled ground truth recorded from a camera.

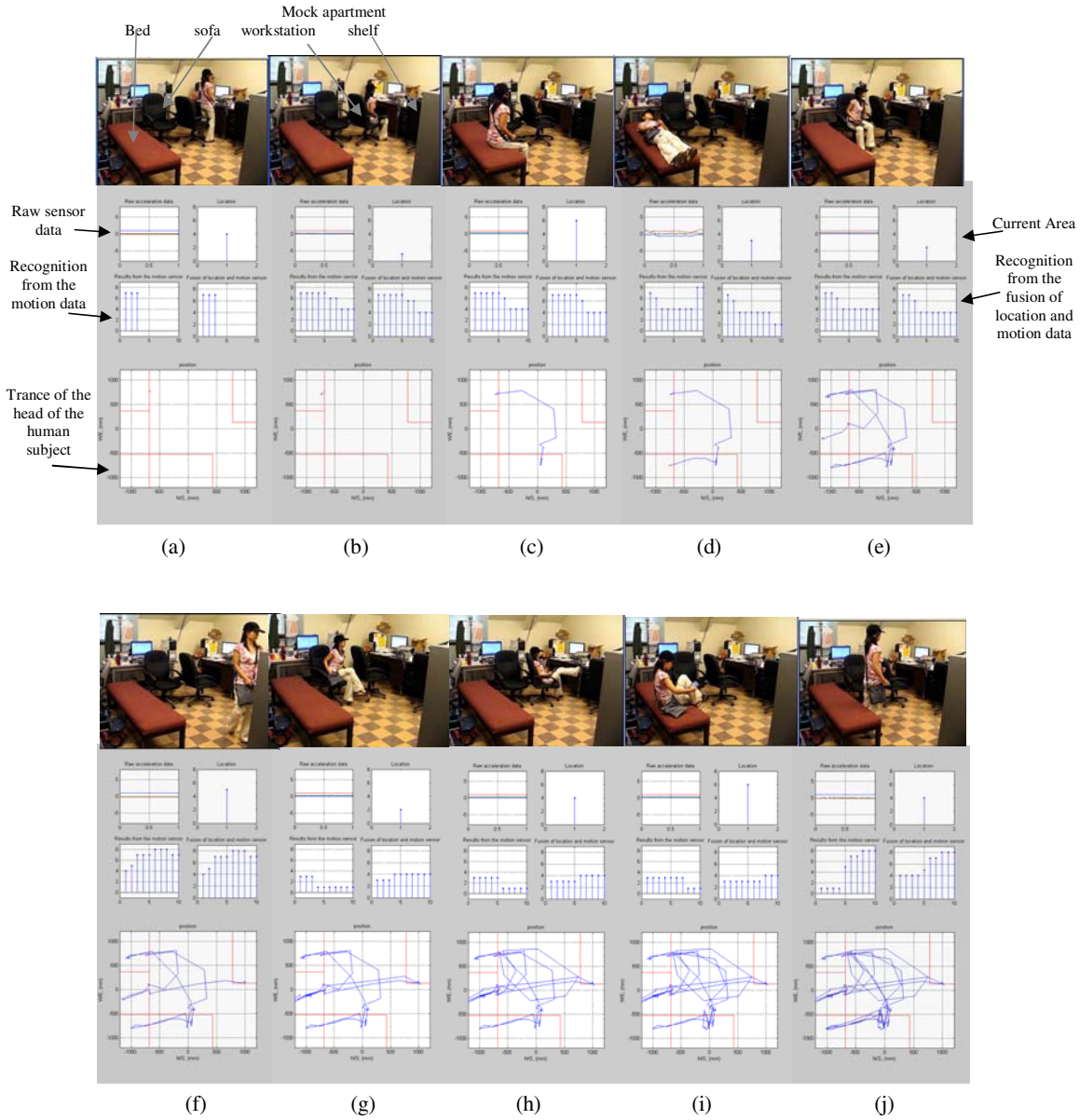

Fig. 7. Snapshots captured from camera and the server PC.

The video of the experiment is synchronized with the figure output of the activity recognition. Some significant frames are shown in Figure 7. From (a) to (j), the top images are from the video of the human subject and the bottom figures are from the server PC screen. In the figure of the recognition results, the top two are the raw sensor data and the segmented location area of the current instant. 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 system. In (a), the human subject starts from standing in location area 4. Both recognition results are the same. In (b), she goes to area 1 and sits down. In (c), she walks to the bed and sits down. In (d), she lies down on the bed. In (e), she sits down onto the sofa. In (f), she walks to the shelf and stands there. In (g), 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 are 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 random movement of leg is not one of the pre-defined activities, it will be recognized as the one with maximum probability after fusing the location

and motion data. The following stationary activity will still be correct because in this area, the probability of sitting is higher than lying. In (h) and (i), she is sitting and moving her leg randomly. Fusion of location can correct the error from lying to sitting. In (j), when 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.

TABLE I
DECISION ACCURACY OBTAINED FROM MOTION DATA ONLY.

Test Type	Decision Type								Test Accuracy
	1	2	3	4	5	6	7	8	
1	**0.80**	0	0	0.20	0	0	0	0	**0.80**
2	0	**0.65**	0.25	0	0	0	0	0.10	**0.65**
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	**0.85**	0	0.05	0.10	**0.85**
6	0	0	0	0	0	**0.81**	0.07	0.12	**0.81**
7	0	0	0	0	0.07	0.03	**0.90**	0	**0.90**
8	0	0	0	0	0	0	0.02	**0.98**	**0.98**

TABLE II
DECISION ACCURACY OBTAINED FROM THE FUSION OF LOCATION AND MOTION DATA.

Test Type	Decision Type								Test Accuracy
	1	2	3	4	5	6	7	8	
1	**0.95**	0	0	0.05	0	0	0	0	**0.95**
2	0	**0.85**	0.15	0	0	0	0	0	**0.85**
3	0	0.10	**0.90**	0	0	0	0	0	**0.90**
4	0.09	0	0	**0.91**	0	0	0	0	**0.91**
5	0	0	0	0	**0.85**	0	0.05	0.10	**0.85**
6	0	0	0	0	0	**0.81**	0.07	0.12	**0.81**
7	0	0	0	0	0.07	0.03	**0.90**	0	**0.90**
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 I and II. 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, fusion of location and motion data can improve the recognition accuracy compared to the recognition using motion data only.

VI. 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 the future, we will combine the location and human activities for simultaneous tracking and activity recognition (STAR) [19], which will remove the need of the OptiTrack motion capture system.

ACKNOWLEDGMENTS

This project is partially supported by the NSF grant CISE/CNS 0916864 and CISE/CNS MRI 0923238.

REFERENCES

[1] C. Zhu and 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*, pages 303–304, 2009.

[2] C. Zhu and W. Sheng. Human daily activity recognition in robot-assisted living using multi-sensor fusion. In *IEEE International Conference on Robotics and Automation*, pages 2154–2159, 2009.

[3] H. A. Yanco and J. L. Drury. Classifying human-robot interaction: An updated taxonomy. In *Proceedings of 2004 IEEE International Conference on Systems, Man and Cybernetics*, pages 2841–2846, 2004.

[4] 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*, pages 90–126, 2006.

[5] C.J. Taylor. Reconstruction of articulated objects from point correspondences in a single image. *Computer Vision and Pattern Recognition*, pages 349–363, 2000.

[6] S. Park and M. M. Trivedi. Multi-person interaction and activity analysis: a synergistic track- and body-level analysis framework. *Machine Vision and Applications*, pages 151 – 166, 2007.

[7] 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 on Biomedical Engineering*, 50:711–723, 2003.

[8] 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.

[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] L. Liao, D. Fox, and H. Kautz. Location-based activity recognition. *Neural Information Processing Systems - NIPS05 Workshops*, 2005.

[11] L. Liao, D. Fox, and H. Kautz. Location-based activity recognition using relational markov networks. *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2005.

[12] B. Lo, L. Atallah, O. Aziz, M. E. ElHew, A. Darzi, and G.Z. Yang. Real-time pervasive monitoring for postoperative care. *4th International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2007)*, pages 122–127, 2007.

[13] U. Maurer, A. Smailagic, D.P.Siewiorek, and 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*, pages 113–116, 2006.

[14] VectorNav Technologies. *http://www.vectornav.com/*. 2010.

[15] Digi International Inc. *http://www.digi.com/*. 2010.

[16] Inc. NaturalPoint. *OptiTrack^{TM} Optical Motion Capture Solutions*. 2009.

[17] J. Bloit and 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*, pages 2121–2124, 2008.

[18] C. Zhu and W. Sheng. Recognizing human daily activity using a single inertial sensor. In *The 8th World Congress on Intelligent Control and Automation*, 2010.

[19] D. Wilson and C. Atkeson. Simultaneous tracking & activity recognition (star) using many anonymous, binary sensors. *Proceedings of PERVASIVE*, pages 62–79, 2005.