LSTM-RNNs Combined with Scene Information for Human Activity Recognition

Wen-Hui Chen Carlos Andrés Betancourt Baca Chih-Hao Tou

Graduate Institute of Automation Technology
National Taipei University of Technology, Taipei, Taiwan
E-mail: whchen@ntut.edu.tw

Abstract*—Developing an accurate activity recognition system with sensor readings is a challenging task due to the fact that sensors are subjected to a variety of errors and the nature of human activities is dynamic and uncertainties. Recent studies have shown that machine learning approaches can effectively classify human activities. In this study, we analyze sensor readings from accelerometers and gyroscopes using long short-term memory (LSTM) recurrent neural networks to identify human activities and present a location-aware approach to improve the recognition accuracy. The location information is obtained from analyzing images captured by a wearable camera based on a pre-trained model of the deep convolutional neural networks. With the location information, some unlikely activities can be ruled out, leading to better recognition accuracy.

Keywords- activity recognition; convolutional neural networks (CNNs); long short-term memory (LSTM); recurrent neural networks (RNNs), senior healthcare

I. Introduction

Thanks to better nutrition, sanitation, and healthcare, people nowadays are living longer. The extended human life span, however, is fostering an aging world. According to the United Nations Population Fund, people aged 60 or more make up 12% of the current global population, and that number is estimated to rise to 22% by 2050 [1].

The increase of aging population with disabilities has become a major concern in developed and developing countries, causing the rising demand for long-term care. The service for home healthcare is an important issue in aging societies worldwide as many reports show that most seniors prefer to age in place. Aging in place refers to the ability to live in one's own home independently.

The decrease of fertility rate in Taiwan has accelerated the progress of aging population, leading to a potential need of senior care. According to the definition of aging society from World Health Organization (WHO), Taiwan has become an aging society since 1993 [2].

The percentage of the population in Taiwan aged over 65 years is predicted to be doubled by 2017 [3]. Additionally, the homecare services in Taiwan are impaired by the shortage of registered nurses and high nurse-patient rate [4]. Therefore, it is necessary to pay more attention to finding alternatives for senior care with the assistance of technology to cope with the rapidly aging population.

Human activity recognition has many potential healthcare applications. Successful examples include helping doctors keep track of patient history and analyze patient behaviors in order to make a better diagnosis. In this study, we endeavor to create an accurate activity recognition system dedicated on activities of daily living (ADLs) using deep learning models that can help develop a senior healthcare system.

The remainder of this paper is organized as follows: Section 2 reviews the related work in activity recognition. Section 3 describes the software framework of the proposed component-based system, followed by experimental results in Section 4. Finally, the conclusions are drawn in Section 5.

II. Related Work

Activity recognition has been investigated for more than two decades. Numerous approaches with different applications have been proposed in past research [5]-[8]. The type of activity recognition can be broadly divided into sensor-based and vision-based recognition.

By deploying sensors in an environment or attaching on a human body, the sensor-based activity recognition attempts to identify different activities from recognizing sensor readings. Although many efforts have been devoted, this type of activity recognition still remains challenging as humans perform activities differently and sensor readings may contain incorrect or noisy data.

As to vision-based activity recognition, the challenging part is to recognize activities through captured images. Therefore, it heavily relies on image and video processing techniques. The privacy issue is another concern in vision-based activity recognition.

Activity recognition can be considered as a classification problem. In the past decade, various classification algorithms have been applied to the design of activity recognition, such as logistic regression, decision trees, naïve Bayes, and support vector machines [9]. Each classification algorithm has its pros and cons, and no single algorithm outperforms others in all cases.

Learning underlying patterns from sensor data is critical for activity inference. Feature extraction is one of

978-1-5090-6704-6/17/$31.00 ©2017 IEEE

important machine learning issues, and also a critical step in activity recognition. Deep learning networks have the ability to perform automatic feature extraction without human intervention. In the past few years, using deep learning networks that avoid relying on hand-crafted features has become a trending research direction and shown promising results in activity recognition [10]-[11].

The location where a person is engaged provides some extent of useful information for activity reasoning. For example, if the subject is in the kitchen, the activity of eating or meal preparation would be more probable than dressing or bathing. When the subject is in a bathroom, it is more probable to take a shower, brush teeth or flush the toilet than take a rest.

In this study, we examine the location information through analyzing images captured from a wearable camera that can be used to increase the accuracy of activity recognition dedicated on ADLs using deep learning models.

III. Software Architecture of the Component-based Activity Recognition System

Considering the software usability and scalability, we proposed component-based software architecture for our system. The proposed framework, as shown in Fig. 1, consists of 10 software components, including Universal Interface (UI), Slow Intelligent System (SIS) Server, System Interface (SI), Alerter, Activity Recognizer (AR), Sensor Processor (SP), Image Processor (IP), Scene Recognizer (SR), Activity Estimator (AE), and Activities of Daily Living (ADL) Knowledge Base.

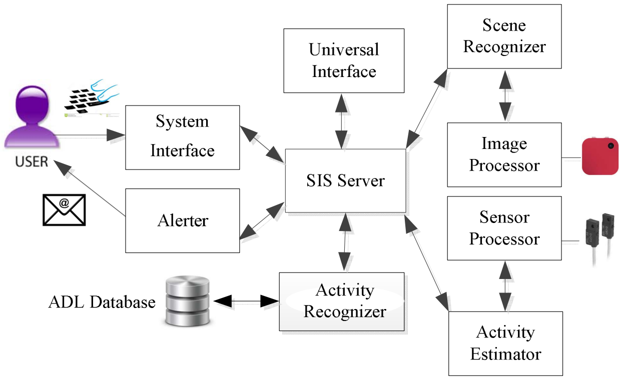

Figure 1. Software architecture of the component-based activity recognition system.

A. Slow Intelligence System (SIS) Framework

We adopted the slow intelligence system (SIS) framework developed by the University of Pittsburgh as the software development platform [12]-[13]. The SIS Server component is responsible for routing and processing messages among components. When the server accepts a request message from a component, the message will be processed and routed to the designated components as responses. All request and response messages are wrapped in the XML format for usability across the Internet. Some particular messages were also defined in the server to execute specific operations.

The **System Interface** (SI) component provides a graphic user interface (GUI) for users to specify system properties, including user names, which e-mail address an alert should be sent to. The properties set by this component are stored in a shared database for other components to access.

The **Universal Interface** (UI) component is used to simulate messages routing among components. Any message sent from one component to another will be displayed. It is useful when testing the developed system by observing routing messages through the UI. The user can enter a message and observe the corresponding output message.

The **Sensor Processor** (SP) component is designed to read, store, and preprocess the raw sensor data for the **Activity Estimator** using recurrent neural networks to classify activities from sensor readings. In addition to analog sensor readings gathered from accelerometers and gyroscopes, the SP can also handle binary sensor data such as pressure switches and proximity sensors deployed in environments.

The **Image Processor** (IP) component is designed to access the images captured by the wearable camera from an egocentric point of view without exposing the user's body to lower the privacy concern.

When an image is received from the Image Processor, the **Scene Recognizer** (SR) component starts to load the image and to detect whether a certain object such as a TV set or bed is present. To identify objects and provide location information, we used the object detection API from deepai.org to detect whether a key object is present. With the detected object as a clue, The Scene Recognizer can infer the possible location of the subject that can help identify activities.

The **Activity Recognizer** (AR) component is designed to recognize activities by the fusion of the estimated activities received from Activity Estimator and the location information received from Scene Recognizer.

The **Alerter** component is responsible for sending an alert e-mail or text messages to notify that an anomaly is detected so that the caregiver can take action accordingly.

B. Recurrent Neural Networks

A traditional neural network lacks of memory elements, so it may not be able to connect previous information to the current task to perform reasoning about previous events. Recurrent neural networks (RNNs) come to address this issue as they use loops in network structure to allow information to persist. This makes RNNs be suitable for processing sequential input data. In the past few years, there has been much success applying RNNs to a variety of problems [14]. Most of those successes are attributed to the use of LSTMs, a special kind of RNNs that are capable of learning long-term dependencies [15]. Therefore, we used LSTM-RNNs to estimate possible activities based on the observation of the sequential sensor data.

There are many notable LSTM variants. A comparison of popular variants can be found in [16]. In this study, we used a basic LSTM model with one input layer, one hidden layer, and one output layer. The number of neurons for input, hidden, and output layers was set to 12, 52, and 13, respectively. The learning rate and the batch size during training were set to 0.0025, 1500, respectively.

IV. Experimental Results

We now turn to present experiments to demonstrate the performance of the proposed approach and examine the effect of location information included for increasing accuracy. For comparison, we also implemented some widely used machine learning approaches as a baseline test.

A. Data Description

To evaluate the effectiveness of the proposed approach, we used the PAMAP2 (Physical Activity Monitoring) dataset that contains data of 18 different physical activities, performed by nine participants wearing three inertial measurement units (IMUs) and a heart rate monitor [17].

The placement of the three IMUs on the human body is as follows: (a) over the wrist on the dominant arm, (b) on the chest and (c) on the dominant side's ankle. In this work, we used only IMU (a) and (b), the one on the wrist and the one on the chest. There are four sensors in each IMU; we used only the sensor readings of one 3D-accelerometer and one 3D-gyroscope, which can be commonly found in wearable devices nowadays. Therefore, the number of features in this work is twelve.

Selecting the length of timeframe and the number of sampling points for each sequential data sample plays an important role in activity recognition. A short timeframe might lead to poor recognition accuracy while having too many sampling points in one sequential sample might be too computational expensive.

The original data was collected at a frequency of 100Hz. Sampling with such a high frequency might not be necessary in consideration of the nature of human activity. Therefore, we set our timeframe to be three seconds and downsized the frequency to 50Hz by extracting only the odd sampling points from the original raw dataset. In other words, of the original 300 sampling points in a timeframe of 3 seconds, we extracted only the 150 odd points, and made it as one sequential data sample.

We modified the dataset based on the criteria mentioned above and separated them into training set and test set at a ratio of 8 to 2 in the stratified random sampling manner. The original dataset included has a total number of 18 activities. However, we did not keep all of them due to the reason that some of the activities do not belong to ADLs. The selected activities along with the number of instances in our experiment are listed in Table 1.

TABLE 1. SELECTED ACTIVITIES FROM PAMAP2 DATASETS

ID	Activities	Training Instances	Test Instances
1	Lying	492	122
2	Sitting	465	116
3	Standing	445	111
4	Walking	153	38
5	Running	113	28
6	Watching TV	220	55
7	Computer work	806	201
8	Ascending stairs	296	74
9	Descending stairs	268	67
10	Vacuum cleaning	420	104
11	Ironing	559	139
12	Folding laundry	240	60
13	House cleaning	455	113
Total		4932	1228

B. Evaluation Criteria

One of the commonly used formulas for classification accuracy is to calculate the number of correct predictions divided by the total number of predictions. As our dataset listed in Table 1 contains unbalanced activity classes, we require a performance metric that is suitable for the biased dataset. Therefore, we used the F1 score as the performance measure to evaluate our work in this study.

Defined in (1), the F-measure or F1-score is a measure of test's accuracy by combining the measure that assesses the precision and recall scores in (2) and (3), respectively.

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{1}$$

$$\text{Precision} = \frac{1}{N} \sum_{i=1}^{N} \frac{TP_i}{TP_i + FP_i} \tag{2}$$

$$\text{Recall} = \frac{1}{N} \sum_{i=1}^{N} \frac{TP_i}{TP_i + FN_i} \tag{3}$$

where
N is the total number of class instances
TP is true positive
FN is false negative
FP is false positive

C. Classification Results

We conducted experiments to evaluate the effectiveness of the LSTM-RNNs on sensor-based activity recognition and compare it with several popular machine learning algorithms, including decision trees, random forests, support vector machines, *k*-nearest neighbors, and naïve Bayes. Comparison results and batch losses during the training phase are shown in Fig. 2 and Fig. 3, respectively. The values of different performance metrics for various classification models are listed in Table 2 for comparisons. From Table 2, we can observe that LSTM-RNNs outperform most widely used classification algorithms.

With the scene information provided by the Scene Recognizer, we can identify the subject's location and rule out unlikely activities. For instance, watching TV is unlikely to take place in a bathroom. Probable activities in certain locations in the experiment are listed in Table 3. Fig. 4 and Fig. 5 illustrate the results where the location was identified as the bathroom and the living room, respectively.

Without location information, the overall classification performance of the LSTM-RNNs in this experiment reaches 82.57%, which outperforms most classification algorithms. Table 4 lists the overall recognition performance without location information. When combining with location information, Activity Recognizer in the proposed system excludes some activities, which improves the activity recognition performance. The overall accuracy when recognizing key objects from captured images and providing location information in the scenario of living room, bathroom, bedroom, and outdoors, the accuracy can be improved at the scale of 83.16%, 88.36%, 82.81%, and 87.56%, respectively.

Fig. 6 shows the effect of location information on recognition performance in the scenario of living room, bathroom, bedroom, and outdoors. As can be observed in Fig. 4, with location information, the overall classification performance can be improved.

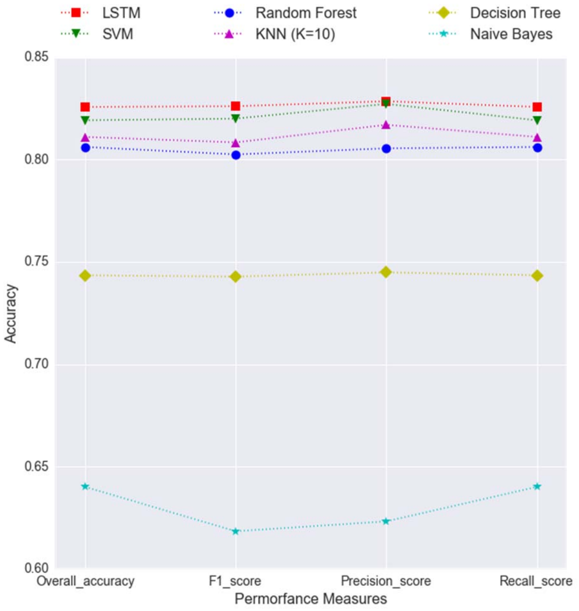

Figure 2. A comparison of recognition performance.

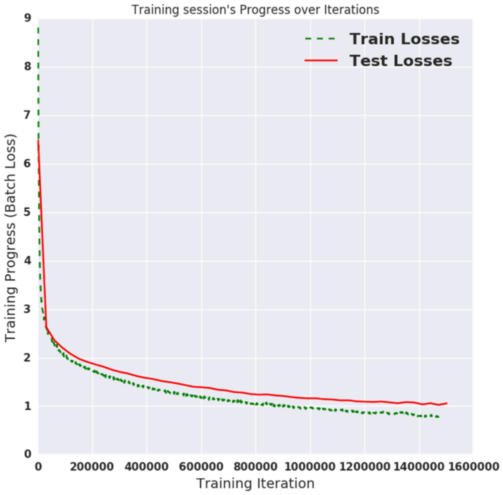

Figure 3. LSTM-RNNs batch losses during training phase.

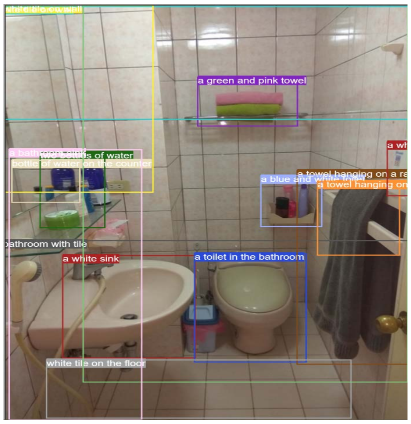

Figure 4. Detection results in a bathroom.

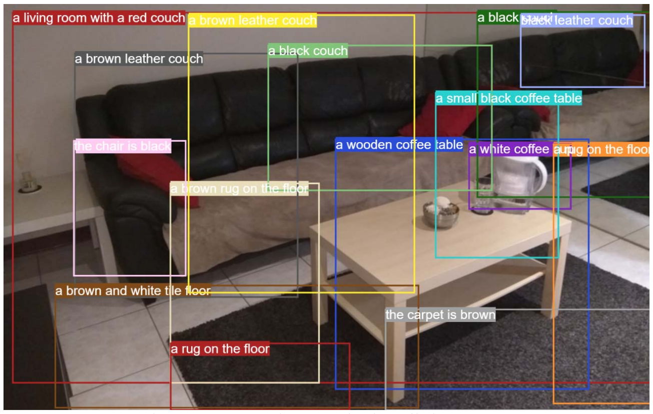

Figure 5. Detection results in a living room.

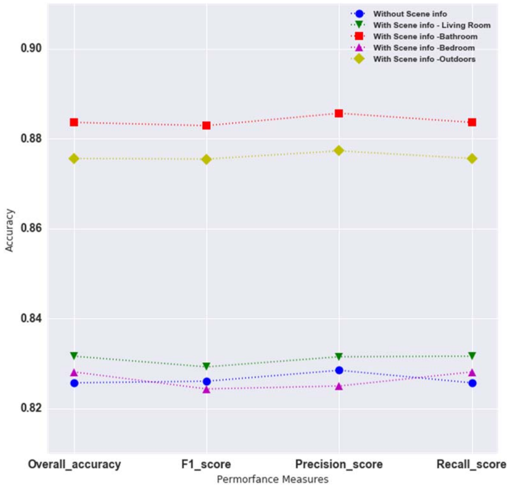

Figure 6. The effect of location information on recognition performance.

TABLE 2. COMPARISIONS OF RECOGNITION PERFORMANCE

Metric / Model	Overall accuracy	F1 score	Precision score	Recall score
LSTM-RNNs	0.826	0.826	0.829	0.826
SVM	0.820	0.820	0.827	0.820
Random Forest	0.806	0.803	0.806	0.806
KNN (k=10)	0.811	0.808	0.817	0.811
Decision Tree	0.743	0.743	0.745	0.743
Naïve Bayes	0.640	0.618	0.623	0.640

TABLE 3. THE RELATIONSHIP OF AREAS AND THEIR PROBABLE ACTIVITIES

Location	**Activities**
Living Room	Lying, Sitting, Standing, Walking, Running, Watching TV, Computer Work, Vacuum Cleaning, Ironing, Folding Laundry, House Cleaning
Bathroom	Sitting, Standing, Walking, House Cleaning
Bedroom	Lying, Sitting, Standing, Walking, Watching TV, Computer Work, Vacuum Cleaning, Ironing, Folding Laundry, House Cleaning
Outdoors	Sitting, Standing, Walking, Running, Ascending Stairs, Descending Stairs

V. Conclusions

Activity recognition has important applications in ubiquitous healthcare. The activity of a senior patient reveals useful information about his or her health status. With activity information, medical professionals can provide senior patients medications efficiently and assess the health problems such as Alzheimer's in an early stage. Activities and locations have some sort of relationships that provide a useful clue for reasoning activities.

In this study, we have presented a software component based framework for activity recognition and deep learning models combined with location information to improve the recognition accuracy. Experimental results have shown that the proposed approach offers a feasible solution to activity recognition for healthcare.

Acknowledgements

This work is supported by Taiwan Ministry of Science and Technology (MOST) under Grant No. MOST 105-2221-E-027-048 -.

TABLE 4. RECOGNITION RESULTS WITHOUT LOCATION INFORMATION

Identifiers **Activities**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Lying	**115**	0	2	1	0	0	0	0	0	1	1	0	2
Sitting	0	**99**	4	0	0	2	3	1	0	0	5	0	2
Standing	0	6	**94**	0	0	0	0	0	2	3	0	2	4
Walking	0	0	2	**28**	1	0	0	6	0	0	0	1	0
Running	0	0	1	1	**21**	0	0	0	0	0	1	2	2
Watching TV	0	1	0	0	0	**53**	0	0	0	0	0	0	1
Computer Work	0	0	1	0	0	0	**197**	0	0	0	0	0	3
Ascending Stairs	0	0	3	5	0	0	1	**62**	2	0	0	0	1
Descending Stairs	0	0	2	0	0	0	0	1	**59**	2	0	0	3
Vacuum Cleaning	0	1	1	0	0	0	0	1	1	**85**	3	4	8
Ironing	0	2	1	0	0	0	0	0	3	1	**105**	11	16
Folding Laundry	0	1	1	0	0	0	1	0	0	2	13	**24**	18
House Cleaning	0	2	4	1	0	0	2	2	4	10	11	5	**72**
F1 Score	97.05%	86.84%	82.82%	75.68%	84.00%	96.36%	97.28%	84.35%	85.51%	81.73%	75.54%	44.04%	58.78%

The number of total test instances: 1228
The number of correct classification: 1014
Overall Accuracy: 82.57%

REFERENCES:

[1] "Ageing in the Twenty-First Century: A Celebration and A Challenge," Published by the United Nations Population Fund (UNFPA), New York, 2012. Available at http://www.unfpa.org/publications/ageing-twenty-first-century (Accessed on May 14, 2017)

[2] J. C. T. Hsueh and Y. T. Wang, "Living Arrangement and the Well-being of the Elderly in Taiwan," International Journal of Welfare for the Aged, vol. 23, pp. 87–107, 2010.

[3] "Taiwan Population Estimation Year 2008 to 2056," Council for Economic Planning and Development, Executive Yuan, Taiwan, 2008.

[4] " The Dark Moment of Nurses in Taiwan," Available: http://ireport.cnn.com/docs/DOC-776292 (Accessed on 10 May 2016)

[5] J. K. Aggarwal and M. S. Ryoo, "Human Activity Analysis: a Review," ACM Computing Surveys, vol. 43, no. 3, pp. 1-47, April 2011.

[6] V. R. Jakkula and D. J. Cook, "Detecting Anomalous Sensor Events in Smart Home Data for Enhancing the Living Experience," Proceedings of the 7th AAAI Conference on Artificial Intelligence and Smarter Living, pp. 33-37, 2011.

[7] W. H. Chen and W. S. Tseng, "A Novel Multi-agent Framework for the Design of Home Automation," Fourth International Conference on Information Technology, pp. 277-281, April, 2007.

[8] W. H. Chen, Y. H. Lin, and S. J. Yang, "A Generic Framework for the Design of Visual-Based Gesture Control Interface," the 5th IEEE Conference on Industrial Electronics and Applications, pp. 1522-1525, June 2010.

[9] O. D. Lara and M. A. Labrador: "A Survey on Human Activity Recognition Using Wearable Sensors," IEEE Communications Surveys & Tutorials, vol. 15, no. 3, pp.1192-1209, 2013.

[10] F. J. Ordo´nez and D. Roggen, "Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition," Sensors, vol. 16, no. 115, pp.1-25, 2016.

[11] N. Y. Hammerla, S. Halloran, and T. Ploetz, "Deep, Convolutional, and Recurrent Models for Human Activity Recognition Using Wearables,"arXiv:1604.08880, 2016.

[12] W. H. Chen, S. K. Chang, and W. P. Hung, "Applications of Slow Intelligence Frameworks for Energy-Saving Control," The 26th International Conference on Software Engineering and Knowledge Engineering, Vancouver, Canada, July 1-3, 2014.

[13] S. K. Chang, W. H. Chen, W. C. Lin, C. L. Thomas, "Application of Slow Intelligence Framework for Smart Pet Care System Design," International Journal of Software Engineering and Knowledge Engineering, vol. 25, 09n10, pp. 1429-1442, Dec. 2015.

[14] B. Hammer, "On the Approximation Capability of Recurrent Neural Networks," Neurocomputing, vol. 31, no. 1-4, pp. 107-123, March 2000.

[15] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, November 1997.

[16] K. Greff et. al., "LSTM: A Search Space Odyssey," arXiv:1503.04069, 2015.

[17] A. Reiss and D. Stricker, "Introducing a New Benchmarked Dataset for Activity Monitoring," The 16th IEEE International Symposium on Wearable Computers, Newcastle, UK, June 2012.