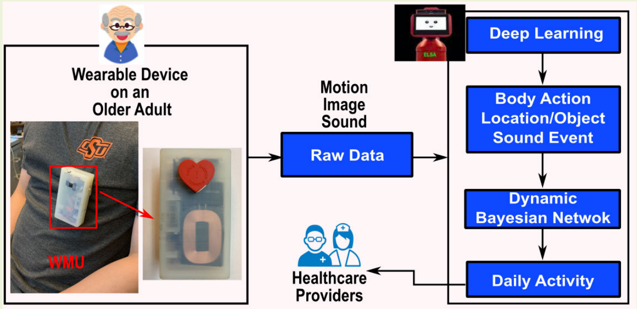
Multimodal Monitoring of Activities of Daily Living for Elderly Care

Fei Liang, *Graduate Student Member, IEEE*, Zhidong Su, *Member, IEEE*, and Weihua Sheng, *Senior Member, IEEE*



***Abstract*—In this article, we presented a multimodal approach to monitor older adults' activities of daily living (ADLs) using the combination of a wearable device and a companion robot. A dynamic Bayesian network (DBN) model was developed for activity recognition, which fuses different data, including location, object, sound event, body action, and time. The walking action is detected as the transition between consecutive activities, which helps capture the inception of activities and save energy on the wearable device. Three tests were conducted to evaluate the proposed approach. First, multiple daily activities were simulated and evaluated the approach based on a public ADL dataset. Second, the proposed approach was tested based on an offline dataset collected in our smart home testbed, which contains images, sound events, motion, and time data. Third, the proposed approach was tested in real time and a web-based interface was developed, which helps caregivers better monitor the ADLs of older adults and provide further assistance. In the offline test and the real-time test, the results show that the system achieved 91% and 93% activity detection ratio, respectively, which significantly outperformed the baseline periodic sampling methods. In addition, the camera and microphone sensor trigger times were reduced from 1537 to 140 and 78, leading to energy reduction of 36.0% and 37.6% on the wearable device, respectively.**

***Index Terms*— Activity monitoring, elderly care, multimodality, wearable computing.**

I. Introduction

THE older adult population around the world is continuously increasing [1]. How to take care of these older adults poses a great challenge. First, old ages are usually associated with many health problems, including physical health deterioration, cognition decline, mental health issues, and so on. Second, aging in place is preferred by a majority of the older adults, as they feel independent and safe in their own homes [2], [3]. The activities of daily living (ADLs) is an important indicator of older adults' well-being [4] and widely used by caregivers in health assessment. ADL monitoring also allows caregivers to provide timely assistance when emergency occurs.

Research in ADL monitoring has been attracting growing interest [5]. Technologies, such as smart homes and the Internet of Things (IoTs) [6], [7], have been widely used in ADL monitoring for elderly care. Cameras, microphones, and passive infrared motion (PIR) sensors were deployed in home environments to monitor ADLs [8]. However, maintaining an infrastructure with many sensors in a home is costly and not practical for many people. In addition, the privacy concern associated with visual sensors, which directly observe and capture user movements, makes people reluctant to accept them, especially for older adults.

On the other hand, mobile or wearable devices, such as smart phones or smart watches, are used by many people in their daily life and can be used to collect data related to ADLs [9]. For example, smart phones are usually equipped with sensors, such as cameras, microphones, and accelerometers, which can capture multimodal data for daily activity recognition [10]. Smart watches can also collect such data for activity recognition, in addition to vital sign data, such as heart rate and blood oxygen level [11]. However, the computation capacity and energy consumption are major concerns for wearables when they have to process large amount of visual and audio data using complicated algorithms, such as deep neural networks. Therefore, it is desirable for wearable devices to collaborate with a more powerful computing resource to accomplish the recognition task. As companion robots can play a major role in elderly care by offering many functions, such as companionship, entertainment, medication reminder, daily

Manuscript received 5 February 2024; accepted 8 February 2024. Date of publication 21 February 2024; date of current version 2 April 2024. This work was supported by the National Science Foundation (NSF) under Grant CISE/IIS 1910993, Grant CPS 2212582, and Grant TI 2329852. The associate editor coordinating the review of this article and approving it for publication was Dr. Ing. Tai Fei. *(Corresponding author: Weihua Sheng.)*

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

Digital Object Identifier 10.1109/JSEN.2024.3366188

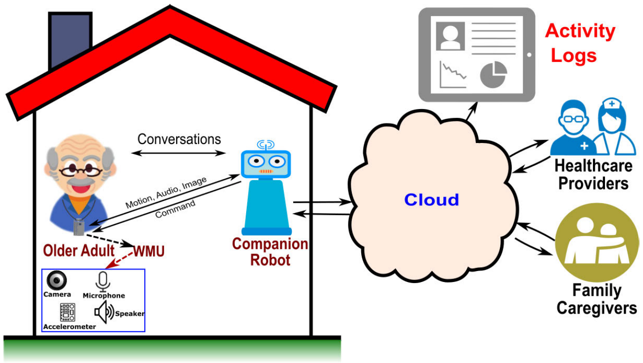


Fig. 1. Concept of MAMS using a companion robot and a wearable device.

life management, and emergency response, we decide to pair the wearable with a robot for collaborative ADL monitoring. The monitoring result can be used by the robot to provide timely intervention to close the loop.

In this article, we mainly focus on the monitoring part of this loop. A multimodal ADL monitoring system (MAMS) was proposed by pairing a wearable device and a companion robot. As shown in Fig. 1, a wearable monitoring unit (WMU) collaborates with a companion robot to collect multimodal data and recognize an older adult's daily activities in real time. Image, audio and motion data are collected by the WMU upon request by the robot. Since the wearable camera only collects the images of the surrounding environment and is only activated when needed, it is less intrusive and more acceptable by older adults in their homes [12]. With the multimodal data, the robot recognizes the corresponding activities by running deep learning algorithms. Meanwhile, the detected activities are logged in a database for further analysis and review by the caregivers and family members.

The contributions of this article are threefold. First, a new dynamic Bayesian network (DBN) model is developed for human activity recognition, which leverages the sequential constraints of ADLs exhibited in the history data of daily life, the activity-location correlation, and the sound associated with the ADL in indoor environments. This multimodal data fusion method ensures the accuracy of ADL recognition. Second, to achieve precise activity segmentation, which facilitates the DBN model in accurately recognizing activities, walking action detection was employed to identify the end of current ADL and the inception of the subsequent ADL. Additionally, it allows the camera and microphone sensors to be activated only when needed instead of periodical sampling, facilitating data collection for activity recognition while conserving battery power on the wearable device. Third, we created a publicly accessible dataset,[1] which consists of image, sound, and motion data collected in our smart home testbed and can be used by the research community working on ADL recognition.

The rest of this article is organized as follows. Section II introduces the related work. Section III presents the overall design of the MAMS. Section IV explains the method and algorithm for energy-efficient and multimodal activity monitoring. Experiments and evaluation results are presented in Section V. Section VI discusses the privacy and scalability issues. Section VII concludes this article and discusses the future work.

II. Related Work

This section surveys the related work in human daily activity recognition, which consists of environmental sensor-based monitoring and wearable sensor-based monitoring. The work regarding energy consumption in wearable sensor-based monitoring was also reviewed.

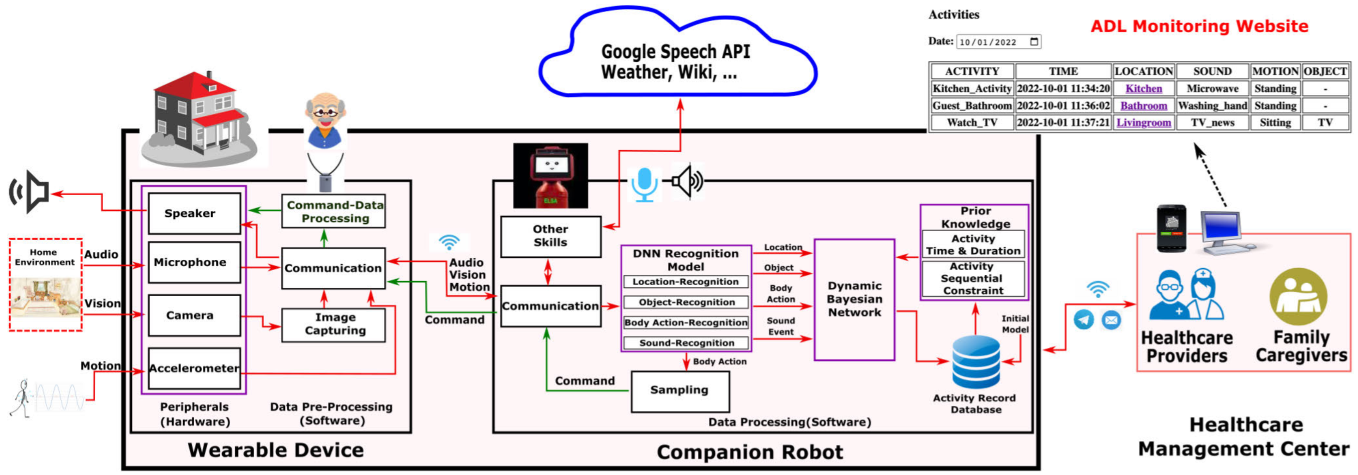
A. Environmental Sensor-Based ADLs Monitoring

Environmental sensors, such as visual sensors, acoustic sensors, and infrared motion sensors, have been used in activity monitoring [13]. Raghav and Chaudhary [14] proposed a fall detection method based on RGB images from surveillance cameras. The authors adopted a deep learning method to recognize a falling person in the images. Eldib et al. [15] developed an ADL monitoring system using low-resolution visual sensors. By deploying ten cameras at different locations in a home, the system could detect 13 daily activities, including eating and cooking. However, one major drawback of video-based ADL monitoring in elderly care is privacy concern [16]. Although depth sensors could be used to alleviate the concern [17], they lack detailed visual information critical to activity recognition. Another issue associated with visual sensors is their limited camera view, which requires multiple cameras to cover the areas of interest.

Acoustic sensors were also used in home environments to detect the daily activities that generate unique sounds. Kim et al. [18] developed a deep learning model for sound-based activity recognition. Their system used a recorder to collect sound data and detect normal activities along with emergency events, such as explosion, glass breaking, and so on. Sim et al. [19] adopted a microphone to collect acoustic data and recognize the corresponding activities. The proposed method could recognize the sounds, such as eating and drinking, with an accuracy of 83.2%. Khan et al. [20] implemented a fall detection system based on sound, which employed two microphones to collect the sound data and recognized the fall sound and nonfall sound using an unsupervised learning method. Due to the limited sensitivity of the microphones and the noise of the environment, audio-based activity monitoring may not work well for the activities that do not generate sufficiently loud sound or when the sound source is far away from the microphones. Although microphones still cause privacy concerns, the devices are more acceptable to people compared to visual sensors [19].

Other sensors were also utilized. In the CASAS project [21], [22], multiple types of sensors, including infrared motion sensors and door switch sensors, were deployed in several apartments to monitor human daily activities. Ghayvat et al. [23] built an activity monitoring system by leveraging distributed sensors, including PIR sensors, electrical object sensor, and other environmental sensors. However, it is expensive and

[1] https://ascclabopensource.github.io



Activities

Date: 10/01/2022

ACTIVITY	TIME	LOCATION	SOUND	MOTION	OBJECT
Kitchen_Activity	2022-10-01 11:34:20	Kitchen	Microwave	Standing	-
Guest_Bathroom	2022-10-01 11:36:02	Bathroom	Washing_hand	Standing	-
Watch_TV	2022-10-01 11:37:21	Livingroom	TV_news	Sitting	TV

Fig. 2. Overall design of the MAMS.

inconvenient to deploy and maintain many distributed sensors in home environments.

B. Wearable Sensor-Based ADLs Monitoring

Wearable sensors have been used to recognize human activities in recent years [9], [10]. For example, an accelerometer was used for the detection of falls in [24] and detection of walking and sitting in [25]. Similarly, a wearable camera was used to detect daily activities, such as cooking and washing hands in home environments [26], [27], which captured the wearer's surroundings through a first-person view. Weiss et al. [28] created an activity monitoring system using a smart phone and a smart watch. By combining motion data from both of them, the system could achieve a high accuracy in activity recognition. However, only a very limited number of activities could be recognized as motion data only are not sufficient to detect many other daily activities, such as reading, watching TV, and so on. Usually single modality-based activity monitoring has low accuracy due to the limited data available. It is generally true that multimodality-based ADL monitoring can achieve better performance and reduce the chance of misclassification [29]. Sun et al. [30] integrated a powerful CPU, a camera, an accelerometer, and an audio processor for dietary monitoring and physical activity monitoring. However, without considering the energy consumption issue, their devices continuously collected data.

Wearable motion sensors can be used to detect transitional activities that are useful to monitor consecutive daily activities [31], [32]. As motion data have small data size and involve less computational complexity, the data can be used to recognize the onset of more complicated activities. For example, Saha et al. [32] utilized the smartphone accelerometer data to recognize the transition activities, including "sit to stand," "stand-walk," and "walk-stand." Okour [33] detected transitional activities based on accelerometer data, which helps recognize different daily activities.

C. Energy Consumption in Wearable-Based Monitoring

Continuous ADL monitoring leads to significant power consumption and reduced battery life of wearable devices. Diyan et al. [34] proposed a duty cycle-based event detection model, in which the authors chose a set of binary motion sensors deployed in a home environment to detect unexpected events. The simulation results showed that the proposed scheme achieved an accuracy of 96.12%, while reducing the energy consumption. However, some events were missed and the selected monitoring sensors consumed a significant amount of battery power, which affected the overall system performance. Possas et al. [35] designed a reinforcement learning-based method to reduce energy consumption of wearable devices by trading off two sensing modalities: vision and motion, as they have different energy consumption and activity recognition accuracy. However, the authors did not study how to handle the short duration activities in daily life. Starliper et al. [36] proposed an activity-aware method to select relevant sensors to reduce the power consumption. This article optimized the sensor set for different types of activities. Then, the proposed method dynamically triggers a selected set of relevant sensors based on the activity the user is doing. The results showed that the proposed method could reduce the power consumption through activity clustering. However, their dataset is small and includes only four activities. As there are many types of activities in daily life, it is difficult to get the target sensor set for each activity in daily life.

Motion sensors can also help reduce the power consumption in ADL monitoring. In our previous work [37], an accelerometer was used to collect motion data continuously and detect potential falls. Upon detection, a wearable camera is triggered to capture image data, which are sent to a robot for fall verification. In this work, we aimed to detect various types of ADLs by leveraging multimodal data including image, audio, and motion data. Particularly, by detecting the transitional walking actions that occur between different activities and leveraging the prediction based on the history of daily activity, the recognition accuracy is improved and energy consumption on the wearable device is reduced.

III. Overview of MAMS

The MAMS consists of a WMU, a companion robot called ASCCBot, and a healthcare management system. The overall design of the MAMS is shown in Fig. 2.

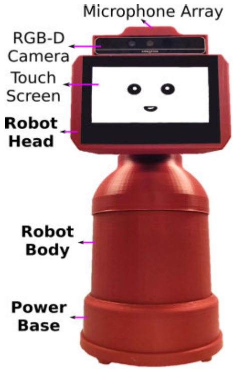


Fig. 3. Prototype of the ASCC companion robot [38].

A. Companion Robot

The custom-designed ASCCBot [38] is the core of the proposed system, which receives data from the WMU and runs the ADL recognition algorithms. As shown in Fig. 3, the robot features a table-top design and runs the neural network models to detect daily activities. The ASCCBot consists of three parts: a head, a body, and a power base. The head contains a mirophone array, an RGB-D camera, and a touch screen display, which is connected to a NanoPi M3 minicomputer running Android OS, while the robot body houses the electronics, which includes a Jetson NX embedded computer running Ubuntu OS. Equipped with natural language processing capabilities, the robot has conversational skills, such as getting weather information, telling jokes, and playing music. It has the ability to communicate with the healthcare center if the older adult needs help. In the MAMS, the robot sends commands to the WMU for data collection and runs the neural network models and the DBN algorithm to recognize the activities based on the data received from the WMU.

B. Wearable Monitoring Unit

As shown in Fig. 4, the WMU has three parts: a main control board, a power module, and a housing. The main control board is built around a Raspberry Pi Zero computer by integrating with a microphone, a bone conduction transducer as a speaker, a Raspberry Pi camera, and a three-axis accelerometer. The WiFi module on the board supports the data transmission between the WMU and the robot.

Through our test, the results showed that when the motion sensor is on to capture data, the average amount of current draw of the circuit is about 255 mA. When the data are transmitted to the robot, the current draw is 380 mA. We incorporated a 2500-mAh lithium-ion polymer battery for the WMU, providing a battery life of approximately 9.55 h. When the device consistently collects data using all the sensors, including motion, camera, and audio sensors, and transmits them to the robot, the battery lasts for 6.58 h. Therefore, it is important to reduce the times of collecting and transmitting visual and audio data. Additionally, a wireless charging module is used, which consists of a receiver coil in the WMU and a transmitter coil outside of the WMU. It takes around 2 h to fully charge the battery as tested. All the sensors and battery are put into a 3-D printed housing, and a strong mounting magnet is used for attaching the WMU on the cloth. In the MAMS, the WMU collects image and audio data based on the commands received from the robot while it continuously collects the motion data to detect the transitional walking actions. The speaker on the WMU allows the older adults to hear what the robots wants to say, such as medication reminders.

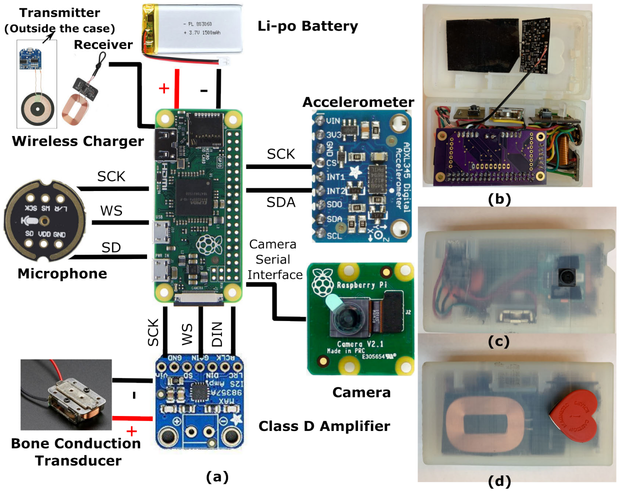


Fig. 4. Design of the WMU. (a) Design of the circuit. (b) Inside. (c) Front. (d) Back.

C. Healthcare Management Center

The healthcare management center logs the daily activity data. Caregivers or family members can review and analyze the ADL data of older adults with authorized access. Meanwhile, the caregivers can offer advice to the older adults to improve their behavioral well-being. When emergency situations occur, such as when a fall is detected, alarms are sent to the caregivers and family members for further help through mobile apps, such as Telegram [39].

D. Working Principle

The overall working principle of the MAMS is shown in Fig. 2. A DBN model is developed to integrate the multimodal sensor data, including location, object, sound event, and motion, while leveraging the sequential constraint, time, and duration learned from the user's history ADL data. As the WMU has limited computational power and battery capacity, the robot and WMU collaborate with each other to accomplish the recognition task. Based on the real-time recognition result, the robot makes decisions on what sensor modalities to use and sends the commands to the WMU. The WMU, upon the request, collects the corresponding sensor data regarding the user and the surrounding environment.

Due to the small data size and low computational cost, the motion sensor runs continuously on the WMU, which helps detect transitional activities like walking. On the other hand, the audio and image data have much larger data sizes and the computation is much more complicated and power hungry. Therefore, the audio and image data are captured and transmitted only upon the request of the robot.

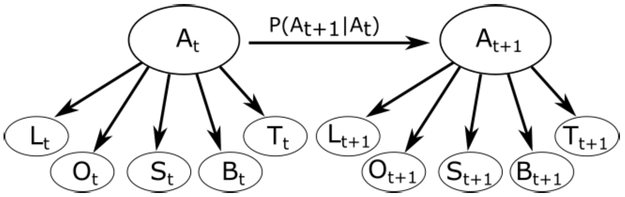


Fig. 5. Graphical representation of DBN model for activity recognition.

IV. Methodology

In this section, the theoretical framework of this research was explained in detail. First the proposed DBN model of ADL recognition was presented. Second, the detection of walking actions was described, which serves as the transition between activities. Third, the different neural network models were discussed, which are used to recognize locations, objects, sound events, and basic body actions for the DBN model. Fourth, the measures for power saving were introduced. Finally, the real-time monitoring system was presented.

A. Dynamic Bayesian Network

In human's daily life, ADLs are highly correlated with the time and the locational contexts, which means particular ADLs usually occur at certain times of the day and particular places in a home. The ADLs also create different types of sound and generate different types of body motion. Such knowledge can be captured by one slice of the DBN model shown in Fig. 5. A_t denotes the activity at time t, L_t denotes the location, O_t denotes the object, S_t denotes the sound event, B_t denotes the basic body action, T_t denotes the time and the activity duration, and P denotes the transition probability from A_t to A_{t+1}. According to Bayes rule, the probability of the predicted activity can be updated by observing the evidences, including locations, objects, sound events, body actions, and time labels, which are collected by the WMU. The time label and the duration of the activity can be achieved based on T_t. As the evidence data L_t, O_t, S_t, B_t, and T_t are dependent on activity A_t, but independent of each other, we have

$$P(A_t|L_t, O_t, S_t, B_t, T_t) = \frac{P(L_t, O_t, S_t, B_t, T_t|A_t) \cdot P(A_t)}{P(L_t, O_t, S_t, B_t, T_t)}. \tag{1}$$

Here, we have the following.

1) $P(A_t| L_t, O_t, S_t, B_t, T_t)$: The posterior probability.
2) $P(L_t, O_t, S_t, B_t, T_t \mid A_t)$: The likelihood function.
3) $P(A_t)$: The prior probability of A_t.
4) $P(L_t, O_t, S_t, B_t, T_t)$: The prior probability of predictor.

Respecting to the independent conditions, we have

$$\begin{aligned} &P(L_t, O_t, S_t, B_t, T_t|A_t) \\ &\quad = P(L_t|A_t) \cdot P(O_t|A_t) \cdot P(S_t|A_t) \cdot P(B_t|A_t) \cdot P(T_t|A_t). \end{aligned} \tag{2}$$

Furthermore, the ADLs also exhibit certain sequential patterns for a particular user. This feature can be utilized to improve the accuracy of ADL recognition, which is represented by the connection between the two slices in the DBN model, as shown in Fig. 5. In this model, the sequences of the activities of daily life form a Markov chain, in which the next activity only depends on the previous one [40], [41]. The hidden states are the daily activities, and the observations include locations, nearby objects, sound events, and body actions recognized by the neural network models, as well as time label and duration, as shown in Fig. 6. The DBN has three sets of parameters: the initial probability distribution, the transition distribution, and the emission probabilities. These distributions are learned from the history labeled activity data by adopting the forward and backward algorithm. For a given sequence of observations, the probability of each activity can be calculated using the forward procedure. Among them, the one with the highest probability is the recognized activity.

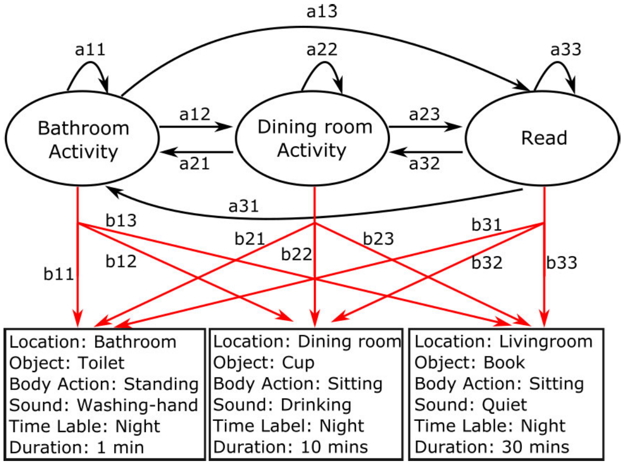
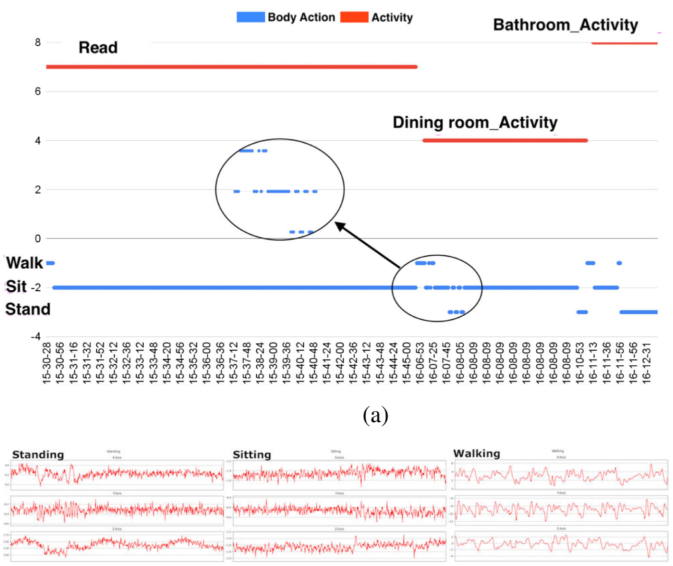


Fig. 6. Example of activities in the DBN.

B. Segmentation of Activities

In the proposed DBN model, it is very critical to correctly segment the different ADLs, i.e., to detect the end of the current ADL and the beginning of the next ADL. In home environments, different ADLs occur at different locations, which means that the transition between two ADLs is typically a walking action. Therefore, detecting the walking action allows us to segment different ADLs. In our proposed method, the motion sensor remains continuously active to detect walking action, which allows the camera and microphone to be triggered, such as at the beginning of the next activity. Fig. 7(a) shows an example. At the beginning, the user sat in the living room reading. Then, he moved to the dining room, where he walked, stood, prepared food, cleaned table, sat, drank water, and had some rest. During one activity, there are still some long walks that could trigger the camera and microphone to collect data. After being recognized by the DBN model, if the recognized activity is the same as the previous one, they are treated as one activity. As only long walks are recognized based on the convolutional neural network (CNN) model described in Section IV-C, the short walks that last less than 2 s are ignored by the model. For example, in the kitchen, walking between the refrigerator and the stoves is ignored. Fig. 7(b) shows the motion data collected by the WMU during standing, sitting, and walking, respectively.



(a)

(b)

Fig. 7. Body actions during activities and motion data samples. (a) Example of motion during the activities. A part of the data is enlarged to show the details. (b) Three-axis motion data.

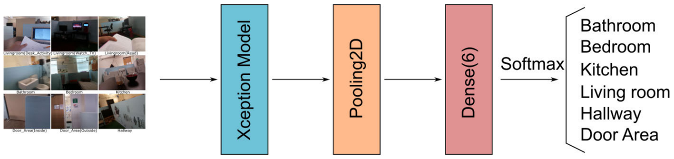


Fig. 8. CNN structure of the location recognition model.

C. Recognition Models for Location, Object, Sound Event, and Body Action

In this section, how to design CNN models to recognize the location, object, sound event, and body action was presented.

1) Location Recognition Model: As shown in Fig. 8, the pretrained Xception [42] model was adopted as the base model to train a neural network for location recognition. The Xception can achieve a top-five validation accuracy of 0.945 and classify 1000 different categories. Then, the final dense layer with a softmax activation function is used to generate the probability distribution of the different locations, which include bathroom, bedroom, kitchen, living rooms, hallway, and door area.

2) Object Recognition Model: YOLO [43] was employed to detect the objects in an image. YOLO is a real-time object detection method based on CNN, and it can detect multiple objects simultaneously and can generate the probabilities for each object in the image. The pretrained model can detect common objects, such as books, TV, laptops, and food items, which is easy to use and meets our object recognition requirement.

3) Sound Event Recognition Model: CNN was used for sound event recognition. The Mel-frequency cepstral coefficient (MFCC) [44] is used as the sound feature. The duration of the sound sample is set to 1 s. The sample rate is 32 000 Hz. The window size of the fast Fourier transform is set to 2048 and the step size is 1024. The number of the Mel band is set to 64. For each sound sample, the generated feature shape is 64 × 32. The CNN network was built, as shown in Fig. 9, which has three convolutional layers (kernel size: 3 × 3 and dimension: 64, 128, and 256) and two dense layers (dimension: 256 and 6). Each convolutional layer is followed by a maxpooling layer (kernel size: 2 × 2). Different numbers of layers and dimensions of the layers were tried and finally chose the abovementioned CNN structure, which has the least parameters and the best performance. The output of each convolutional layer is processed by a batch normalization and a rectified linear unit (ReLU) activation function.

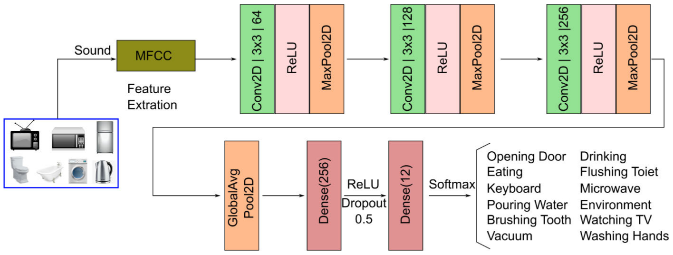


Fig. 9. CNN structure of the sound event recognition model.

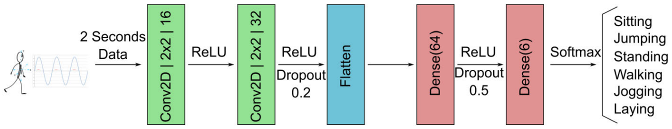


Fig. 10. CNN structure of the body action recognition model.

4) Body Action Recognition Model: As shown in Fig. 10, a CNN model was built for walking action recognition. The model extracts features from one sample with 2 s of acceleration data [45]. It contains two convolutional layers and two dense layers. The convolutional layers have a kernel size of 2 × 2, and the dimensions are 16 and 32, respectively. The dense layers are used to learn features from the combined features of the previous layer. During the training, the Adam algorithm [46] was used for optimization and dropout for overfitting prevention.

D. Measures for Power Saving

As the WMU runs on battery power, it is very important to reduce its power consumption during ADL monitoring. The WMU has a Raspberry Pi Zero as the main processor, which has very limited computational power. In order to test its performance, the CNN models were converted into the Tensor Lite models and deployed them on the WMU. However, it costs an extra 240 mA of current draw to run the models, which is too much, considering that there are four CNN models to run. On the other hand, it only takes 2.225 s and an extra 222.5-mA current draw to send all the data to the robot for recognition. A similar scenario occurs in sound event recognition. Hence, it is more energy efficient for the WMU to collect data and send them to the robot for processing. The robot is capable of data analysis, including recognizing the activities by running the DBN model, recognizing locations, objects, sound events, and body actions, generating the sampling requests, and running the real-time monitoring system.

As mentioned in Section II-C, continuous ADL monitoring takes a significant amount of energy. Some approaches adopt

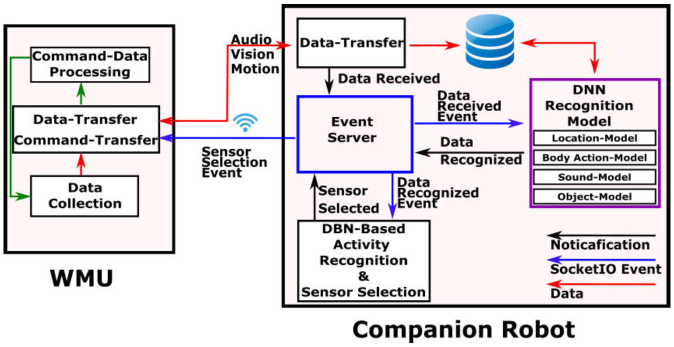


Fig. 11. Workflow of the real-time communication.

TABLE I
DATASET OF LOCATIONS AND BODY ACTIONS

Location			Body Actions	
Location	Samples		Body Action	Samples
Bathroom	332		Sitting	260
Bedroom	864		Jumping	260
Kitchen	1207		Standing	260
Living room	1375		Walking	260
Hallway	280		Jogging	260
Door Area	430		Laying	260

TABLE II
DATASET OF SOUND EVENT

Index	Event	Samples		Index	Event	Samples
0	Opening_Closing Door	99		6	Drinking	118
1	Eating	114		7	Flushing Toiet	87
2	Keyboard	114		8	Microwave	116
3	Pouring Water	125		9	Environment	200
4	Brushing Tooth	123		10	Watching TV	129
5	Vacuum	130		11	Washing Hands	103

duty cycle methods to periodically collect data for activity recognition. However, selecting the duty cycle threshold poses a challenge. A small cycle consumes more energy, while a large cycle may result in missing activities. In our study, we propose to keep the motion sensor on, which is used to identify walking actions as transitional motions between two activities. The walking actions can trigger both the camera and microphone to collect data. In this way, the camera and microphone trigger times can be reduced and more activities can be detected compared to the periodic sampling methods.

E. Real-Time ADLs Monitoring

From Fig. 2, we can see that in the system, the WMU collects and sends data to the robot for activity recognition. The data and the corresponding results are logged in the database. As shown in Fig. 11, an event server was designed for the real-time communication between the WMU and the robot based on SocketIO [47]. The WMU client module keeps the connection with the event server and listens to the sensor selection event (command) to turn on the corresponding sensors and collect and send data to the data receiver module on the robot. The event server publishes the *data received* event to the DNN recognition module for data analysis. After finishing the recognition tasks, the DNN recognition module notifies the event server to publish the *data recognized* event. Then, the DBN model generates the activity result and pushes the sensor selection result to the event server. Finally, the event server pushes the *sensor selection* event to the WMU for data collection.

V. EXPERIMENTAL EVALUATION

In this section, we first demonstrate the performance of the CNN models for location recognition, sound event recognition, and body action recognition; then, three test scenarios were proposed to evaluate the proposed DBN approach for ADLs monitoring based on a public ADL dataset, an offline dataset, and a real-time test, respectively. Finally, the user interface of the real-time ADL monitoring system was demonstrated.

A. Evaluation of Individual Recognition Modules

1) Test Setup: The three proposed CNN models and YOLO V3 [43] were implemented for location, sound event, motion, and object recognition on a PC with a 16-core Intel i9 CPU and an Nvidia Geforce RTX 3070 GPU. The models work with Python 3.7 and Tensorflow 2.8.0. For the location recognition and body action recognition model, as shown in Table I, the dataset contains six locations and six body actions. For sound event recognition, as listed in Table II, a total of 12 events are considered. One subject wore the WMU to collect data ten times in our lab. Location images were collected by entering different rooms to capture diverse data. For sound data, the prerecorded audio was utilized to simulate sound events, segmenting the audio data into 1-s intervals for training and evaluation purposes. For motion data, we followed the actions outlined in Table I and continuously collected data for 10 min, subsequently dividing the data into 2-s segments for each action. During model training, the dataset was partitioned into training, testing, and validation sets using a ratio of 3:1:1. For the location recognition model, the Adam optimizer was utilized, with the epoch set to 10, batch size to 128, and learning rate to 0.001. For the motion recognition model, the Adam optimizer was utilized, with the epoch set to 700, the batch size to 128, and learning rate to 0.0001. For the sound event recognition model, the stochastic gradient descent (SGD) optimizer was utilized, with the epoch set to 70, the batch size to 128, and learning rate to 0.01.

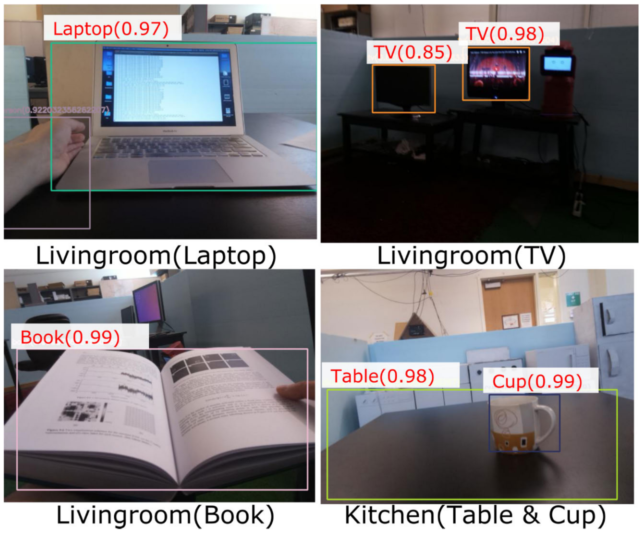
2) Results and Analysis: Following K-fold cross-validation with k set to 5, the proposed recognition models were evaluated. The location recognition model achieved an overall accuracy of 97.39%. The confusion matrix is shown in Table III. The precision, recall, and F1 score are 0.9841, 0.9794, and 0.9793, respectively. The bedroom location sometimes is recognized as a hallway or a door, since the wall of the bedroom is similar to the wall of the hall, the door area, and the living room. Similarly, the overall accuracy of body action recognition is 95.23% and the confusion matrix is shown in Table IV. The precision, recall, and F1 score are 0.9533, 0.9536, and 0.9535, respectively. Particularly, the walking action can be recognized with high accuracy, which is important to detect the next activities. Furthermore, the overall accuracy of sound event recognition is 91.56%, and Table V shows the confusion matrix. The precision, recall, and F1 score are 0.9120, 0.9332, and 0.9225, respectively. Generally, the model can detect with high accuracy all the events,

TABLE III
CONFUSION MATRIX OF LOCATION RECOGNITION

		Predicted Classes					
		Bathroom	Bedroom	Kitchen	Living Room	Hallway	Door Area
Real Classes	Bathroom	1.00	0.00	0.00	0.00	0.00	0.00
	Bedroom	0.01	0.94	0.02	0.03	0.00	0.00
	Kitchen	0.00	0.00	1.00	0.00	0.00	0.00
	Living Room	0.00	0.00	0.03	0.97	0.00	0.00
	Hallway	0.00	0.00	0.00	0.00	1.00	0.00
	Door Area	0.00	0.00	0.00	0.00	0.00	1.00

TABLE IV
CONFUSION MATRIX OF BODY ACTION RECOGNITION

		Predicted Classes					
		Jogging	Jumping	Laying	Sitting	Standing	Walking
Real Classes	Jogging	1.00	0.00	0.00	0.00	0.00	0.00
	Jumping	0.07	0.93	0.00	0.00	0.00	0.00
	Laying	0.00	0.00	1.00	0.00	0.00	0.00
	Sitting	0.00	0.00	0.00	0.93	0.04	0.03
	Standing	0.00	0.00	0.00	0.11	0.89	0.00
	Walking	0.00	0.00	0.00	0.00	0.02	0.98



Livingroom(Laptop) Livingroom(TV) Livingroom(Book) Kitchen(Table & Cup)

Fig. 12. Samples of object recognition result.

including *watching TV, vacuuming, drinking*, and *environment (background sound)*, which is helpful to distinguish the similar activities occurring at the same location. Fig. 12 shows the object recognition results using YOLO and OpenCV. We can see that the objects, such as laptop, TV, and book, can be detected accurately, which gives more information to recognize different activities that occur at the same location. For object recognition, a confidence threshold of 90% is applied, and we focus on the results exceeding this threshold.

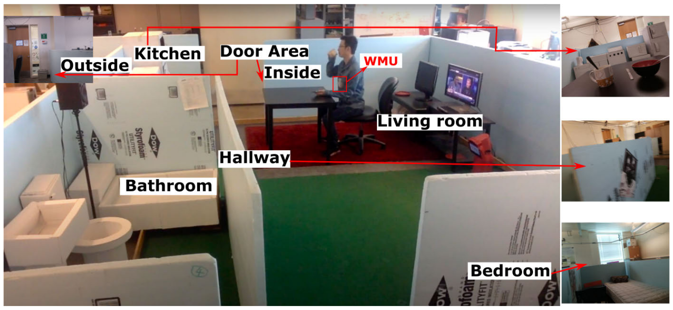
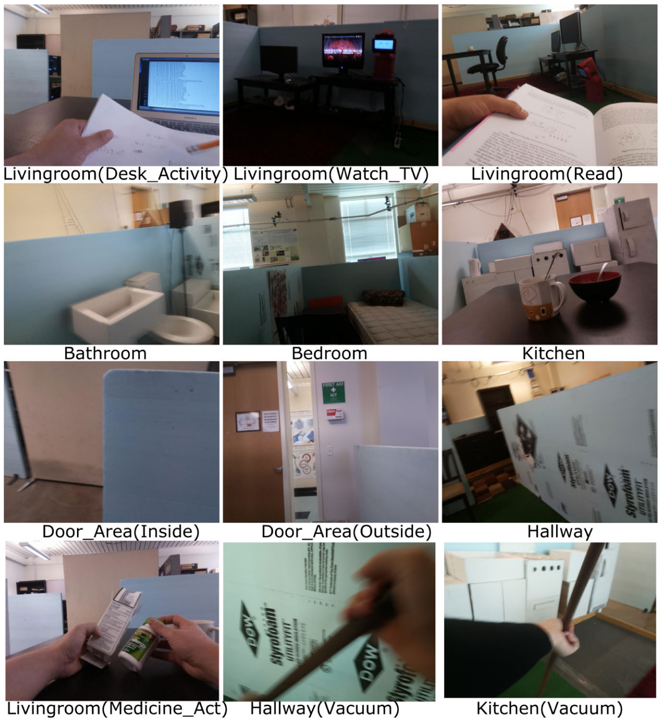


Fig. 13. ADL monitoring system testbed.



Livingroom(Desk_Activity) Livingroom(Watch_TV) Livingroom(Read) Bathroom Bedroom Kitchen Door_Area(Inside) Door_Area(Outside) Hallway Livingroom(Medicine_Act) Hallway(Vacuum) Kitchen(Vacuum)

Fig. 14. Samples of locations and activities from offline dataset.

B. Performance for ADL Monitoring

1) Test Setup: To evaluate the proposed approach, first, the experiments were conducted based on the data from the smart home CASAS project [48] to learn the transition probabilities of the DBN. Milan is selected, which contains 14 activities performed over a span of three months in a smart home, where sensors are deployed at different locations. The activities are shown in Table VI. As some activities' minimum duration is within 0.2 min, it is hard to detect with a periodic method. Second, as shown in Fig. 13, a subject worn the WMU and created an offline dataset which includes image, sound and motion data in our ASCC smart home testbed, by following the activity routine of the date 12/11/2009 from Milan dataset. Then, the proposed system was evaluated on the offline dataset. Fig. 14 presents some sample scenes of the routine from the offline dataset. Finally, a real-time test was conducted by wearing the WMU and following the daily pattern. The maximum activity duration is limited to 5 min to shorten the test time. Additionally, a website was designed to display the detected activities and the corresponding locations, objects, sound events, and body actions information.

2) Simulation Results and Analysis: First, the CASAS dataset history data were split into a training set and a test set with the ratio 3:1 to evaluate the performance of the DBN model, and the model could predict the next activities accurately with an accuracy of 92.8%. Furthermore, the probability distribution of the four relevant activities was listed in Fig. 15. Particularly, several activities that occur at the same location usually have higher probabilities compared with others when the system detects that location. For example, as a person usually does the *Desk_Activity* and *Read* activity in the living room with a sitting action, the system detects these two

TABLE V
CONFUSION MATRIX OF SOUND EVENT RECOGNITION

		Predicted Classes											
		Opening Closing Door	Eating	Keyboard	Pouring Water	Brushing Tooth	Vacuum	Drinking	Flushing Toiet	Microwave	Environment	Watching TV	Washing Hands
Real Classes	Opening Closing Door	0.78	0.04	0.04	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.11
	Eating	0.00	1.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00
	Keyboard	0.00	0.21	0.79	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Pouring Water	0.00	0.18	0.00	0.82	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Brushing Tooth	0.00	0.05	0.00	0.00	0.95	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Vacuum	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	Drinking	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
	Flushing Toiet	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.93	0.00	0.00	0.00	0.07
	Microwave	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.94	0.06	0.00	0.00
	Environment	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.96	0.00	0.00
	Watching TV	0.04	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.85	0.00
	Washing Hands	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.93

TABLE VI
ACTIVITY INDEX AND DURATION FROM THE MILAN DATASET

Activity	Index	Min	Mean	Max
Bed-to-Toilet	1	0.5	0.9	6.2
Morning_Meds	2	0.2	1.0	4.4
Watch_TV	3	2.1	34.3	154.3
Kitchen_Activity	4	0.2	12.3	107.2
Chores(Vacuum Cleaning)	5	2.4	26.3	74.7
Leave_Home	6	0.2	19.7	154.2
Read	7	1.5	23.8	123.0
Guest_Bathroom	8	0.2	2.1	16.1
Master_Bathroom	9	0.2	4.9	45.1
Desk Activity	10	0.5	10.8	52.8
Eve_Meds	11	0.2	0.5	2.1
Meditate	12	1.5	6.4	14.9
Dining_Rm_Activity	13	2.35	12.2	36.7
Master_Bedroom_Activity	14	0.2	18.6	85.2
Note: The Duration is with minute				

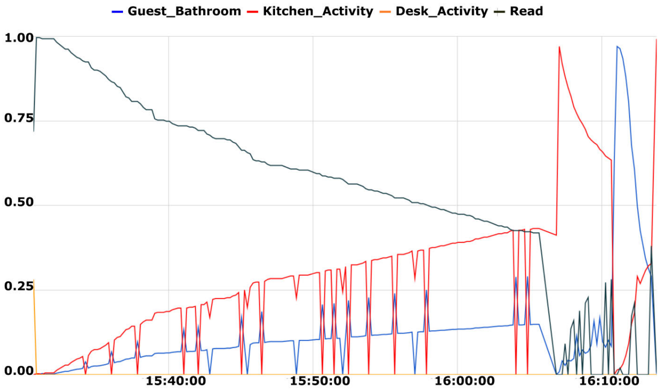
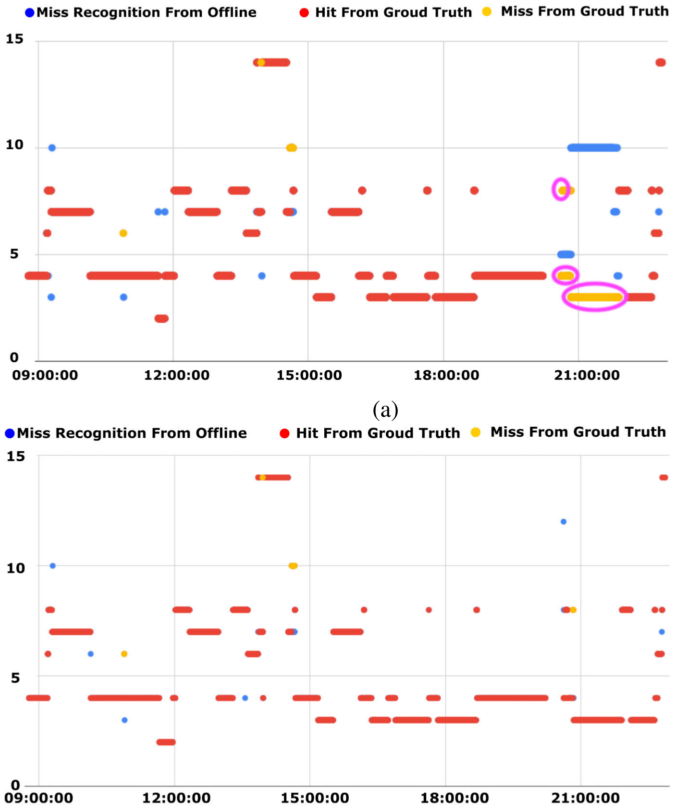


Fig. 15. Probability distribution of four activities during 15:30:39–16:13:47.



(a)

(b)

Fig. 16. Distribution of activities during one day. (a) Activities distribution based on vision and motion data. (b) Activities distribution based on vision, sound, and motion data.

activities with high probabilities at the beginning of the *Read* activity. Meanwhile, the system uses the object information to further confirm the *Read* activity since a book is recognized in the scene. After that, as the duration of *Read* activity increases, the probability decreases. On the other hand, the probabilities of other activities that may occur there increase, such as *Kitchen* activity and *Guest_Bathroom* activity. When the system detects the walking action and recognizes the corresponding new locations, objects, and sound events, the new activity is recognized as *Kitchen* activity with a high probability by the DBN.

3) Offline Evaluation Results and Analysis: Based on the offline dataset, two scenarios were compared: 1) using two types of data: vision and motion data and 2) using three types of data: vision, motion, and sound data. From Fig. 16a, we can see the activity distribution over a span of one day when using the vision and motion data. According to the accuracy definition in [34], the activity detection ratio is 85.1%, with the system successfully identifying 40 out of 47 activities. For the seven missing activities (shown as yellow dots), the misdetection of the walking action leads to three missed activities, including *Leave_Home* (10:53:06), *Master_Bedroom* (13:56:52), and *Desk_Activity* (14:35:32). For example, in the *Leave_Home* (10:53:06) activity, the user walked quickly to open the door, walked to the trash can, and then returned to the home without stop, which makes it hard to detect the end of the walking, while the location information was not available as the camera was not triggered. On the other hand, due to the lack of the sound event information, the other four activities are missed. For example, the system misclassified the three activities, including *Kitchen* activity (20:36:42), *Guest_Bathroom* (20:38:44), and *Guest_Bathroom* (20:49:02) as *Chores (Vacuum Cleaning)* activity between the time of 20:36:42 and 20:49:47. There are two reasons: first, the DBN predicted the *Chores (Vacuum Cleaning)* activity with a high probability as the user usually does the chores everyday; second, it needs more information to distinguish similar activities. In this case, the *Kitchen* activity and the *Vacuum Cleaning* activity can occur in the kitchen. The *Guest_Bathroom* activity and the *Vacuum Cleaning* could both happen in the bathroom. Meanwhile, the human motion is similar as the user can sit, walk, and stand, which made it hard to distinguish the activities only relying on vision and motion data. On the other hand, if the images are blurred, sound information is helpful to confirm the activities. For example, at 20:50:19, the *Desk_Activity* was recognized as *Watching_TV* due to the blurred images captured by the WMU, which calls for sound information for correct recognition. Fig. 16(b) shows that with the help of sound event, several similar activities were correctly distinguished. Compared with the activities labeled with the purple circles in Fig. 16(a), the *Kitchen* activity (20:36:42), *Guest_Bathroom* (20:38:44), and *Watching_TV* (20:50:19) were recognized correctly in Fig. 16(b). Furthermore, there are less blue dots, which means with the help of sound event, the system could recognize the activities more accurately. It could detect 43 out of 47 activities. However, if the activities have a short duration or the walking motion is not detected, the proposed system would miss the activities. Additionally, some incorrect recognition results occur at the beginning of a new activity as the small blue dots shown in Fig. 16(a) and (b), but they could be corrected quickly.

4) Evaluation of Real-Time ADL Monitoring: Fig. 17 shows the results of a real-time test, which achieve good performance and only miss three activities, including two short activities, i.e., the *Leave Home* and the *Master Bedroom* activity. The *Read* activity was recognized as *Watching TV*. This is because the camera images do not contain the book but the TV instead. As the sound event is *Environment*, it cannot be used to distinguish the *Watching TV* and the *Read* activities in the living room. In real-time testing, the detection ratio remains comparable to that of offline testing, owing to the dynamic environmental shifts. At times, real-time testing even demonstrates better performance, as shown in Table VII.

Figs. 18 and 19 show the ADL monitoring system website, which records the historical ADLs and displays the activities and the corresponding locations, sound events, body actions, and objects based on the date the user selects. Users can view the location source images by clicking the hyperlink. This website can help caregivers to better understand the daily activities of the older adults.

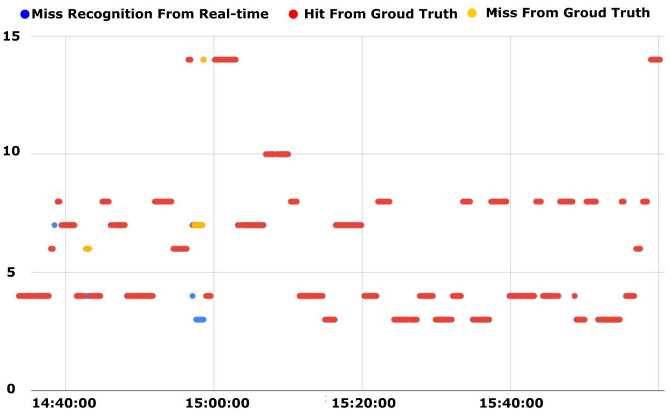
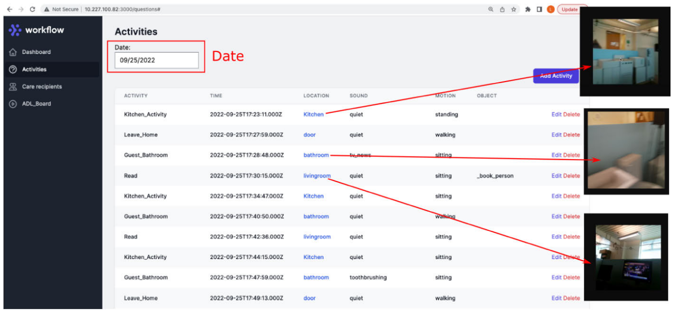


Fig. 17. Distribution of activities in a real-time test.

TABLE VII
RESULTS BETWEEN PERIODIC METHODS AND THE PROPOSED METHOD

Method	Detection Ratio	Trigger Times	Battery Life (h)
0.5 Mins(Period)	89%	1537	7.02
1 Mins(Period)	85%	805	8.35
2 Mins(Period)	70%	412	9.06
3 Mins(Period)	68%	277	9.30
Proposed_vision+motion	85%	276	9.30
Proposed_vision+motion+sound	91%	140	9.55
Proposed_real_time_test	93%	78	9.66



Activities

Date: 09/25/2022

ACTIVITY	TIME	LOCATION	SOUND	MOTION	OBJECT	
Kitchen_Activity	2022-09-25T17:23:11.000Z	Kitchen	quiet	standing		Edit Delete
Leave_Home	2022-09-25T17:27:59.000Z	door	quiet	walking		Edit Delete
Guest_Bathroom	2022-09-25T17:28:48.000Z	bathroom	tv_news	sitting		Edit Delete
Read	2022-09-25T17:30:15.000Z	livingroom	quiet	sitting	_book_person	Edit Delete
Kitchen_Activity	2022-09-25T17:34:47.000Z	Kitchen	quiet	sitting		Edit Delete
Guest_Bathroom	2022-09-25T17:40:50.000Z	bathroom	quiet	walking		Edit Delete
Read	2022-09-25T17:42:36.000Z	livingroom	quiet	sitting		Edit Delete
Kitchen_Activity	2022-09-25T17:44:15.000Z	Kitchen	quiet	sitting		Edit Delete
Guest_Bathroom	2022-09-25T17:47:59.000Z	bathroom	toothbrushing	sitting		Edit Delete
Leave_Home	2022-09-25T17:49:13.000Z	door	quiet	walking		Edit Delete

Fig. 18. Web page showing activities during the day of 09/25.

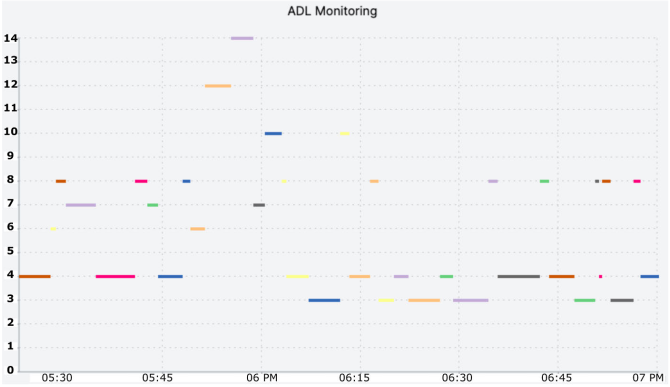


Fig. 19. Web page showing the activity distribution during the day of 09/25.

5) Evaluation of Energy Consumption: To evaluate the energy consumption, we compared the results from the proposed method with those from a baseline periodic method, in which the sensors are periodically turned on to collect data. The trigger times are used as the measure of energy cost since it is directly related to the power consumption on the wearable device. Based on the measurements, it costs extra

0.45 W and 1.79 s to send the image and motion data to the robot. When combined with audio data, it takes 2.225 s. As shown in Table VII, when the period increases, the trigger time decreases, but the detection ratio decreases as well. With the help of the transition motion information, the proposed method triggered 140 times and achieved a detection ratio of 91%, which reduced the sensor trigger times by 90.9% compared with the 0.5-min periodic triggering method. During real-time testing, there is a 94.9% reduction in sensor trigger times. When considering about the overall battery life, it saves 36.0%.

We also compared the proposed method with the methods in two other papers. Diyan et al. [34] utilized environmental sensors to monitor human daily activities, achieving a high activity recognition accuracy of 96.12%. However, their approach comes with a higher cost, and the energy consumption depends on the number of sensors deployed. On the other hand, Sun et al. [30] developed a wearable device utilizing a periodic data collection method with a 2-s interval for activity recognition. It is energy-consuming. Using their method, the device's battery can last only 3 h when tested in on our scenario. Our method incorporates context information related to walking actions and employs a multimodal approach for monitoring daily activities. This approach is designed to conserve energy while delivering robust performance.

VI. Discussion

A. Privacy

Older adults are sensitive to their privacy [16]. Therefore, it is very important to consider privacy protection in the system design. In the proposed monitoring system, three types of data are used.

1) The motion data collected by the accelerometer are generally not considered to be sensitive.
2) The audio data collected by the microphone are more acceptable to the older adults compared to the surveillance camera, as evidenced by the wide adoption of smart speakers like Amazon Echo in many homes. In addition, we activate the microphone only when needed, therefore avoiding continuous sound data collection.
3) The image data collected by the camera are from a first-person perspective and only capture the surrounding environment. Similar to the microphone, it is only activated occasionally, which further mitigates the privacy concern.

In addition, data security can enhance privacy protection [49]. In our system, the raw data are stored locally on the robot, and the access is restricted to authorized caregivers. When high level data are accessed by healthcare professionals and family members through the Internet, security features, such as encryption and password, are used to provide additional protection.

B. Scalability

This article evaluates the system in a laboratory environment, adhering to a specific daily routine. For real-world deployment, it is imperative to gather a large dataset to thoroughly evaluate and enhance the system's performance. To accommodate new environments, the system, equipped with stored daily life data, can involve caregivers or family members in labeling unknown activities. Subsequently, the model can be retrained to adapt to these new environments. In addition, we are currently working with a professor in the College of Human Sciences and Education who has an apartment available for us to test the system in a realistic home environment.

VII. Conclusion and Future Work

In this article, a multimodal approach was designed to monitor older adults' ADLs through the collaboration of a wearable device and a companion robot. First, a DBN model was developed for activity recognition, which fuses different data, including location, object, sound event, motion, and time. Second, the walking action is detected as the transition between consecutive activities, which helps capture the beginning of the next activity and save energy on the wearable device. To test the proposed system, three scenarios were used to evaluate the proposed DBN approach for ADLs monitoring: the public ADL dataset, offline dataset, and real-time test. A real-time monitoring system was developed to help caregivers better understand the daily life of older adults and provide further assistance. The results show that, compared to baseline methods, our method has better accuracy by detecting 43 out of 47 activities in the offline test and 44 out of 47 activities in the real-time test. The sound and object information is useful to distinguish the similar activities that occur at the same locations. Furthermore, by treating the walking actions as the transition motion, the proposed method significantly reduced the sensor trigger times compared with the 0.5-min periodic triggering method. The overall battery life was increased by 36.0% and 37.6% in the offline and real-time test, respectively. However, the proposed system still has some limitations. In the future work, we will improve the system in the following ways: 1) developing activity recognition models that would run on the WMU and studying the tradeoff among energy consumption, recognition accuracy, and time cost in collaborative monitoring; 2) investigating the significance of each data modality and enhancing the system's performance; 3) leveraging the microphones and cameras on the robot for data collection in addition to the WMU sensors; 4) developing robot intervention capabilities to close the loop for the robot to assist older adults when there are behavioral anomalies; and 5) evaluating the proposed algorithms and system in real life environments.

References

[1] *Ageing and Health*. Accessed: Feb. 16, 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health

[2] J. L. Wiles, A. Leibing, N. Guberman, J. Reeve, and R. E. S. Allen, "The meaning of 'aging in place' to older people," *Gerontologist*, vol. 52, no. 3, pp. 357–366, Jun. 2012.

[3] M. J. Bárrios, R. Marques, and A. A. Fernandes, "Aging with health: Aging in place strategies of a Portuguese population aged 65 years or older," *Revista de Saúde Pública*, vol. 54, pp. 1–11, Dec. 2020.

[4] M. Hopman-Rock, H. van Hirtum, P. de Vreede, and E. Freiberger, "Activities of daily living in older community-dwelling persons: A systematic review of psychometric properties of instruments," *Aging Clin. Experim. Res.*, vol. 31, no. 7, pp. 917–925, Jul. 2019.

[5] N. Camp et al., "Technology used to recognize activities of daily living in community-dwelling older adults," *Int. J. Environ. Res. Public Health*, vol. 18, no. 1, p. 163, 2021.

[6] A. Zielonka, M. Wozniak, S. Garg, G. Kaddoum, Md. J. Piran, and G. Muhammad, "Smart homes: How much will they support us? A research on recent trends and advances," *IEEE Access*, vol. 9, pp. 26388–26419, 2021.

[7] O. Djumanazarov, A. Väänänen, K. Haataja, and P. Toivanen, "An overview of iot-based architecture model for smart home systems," in *Proc. Int. Conf. Intell. Syst. Design Appl.* Cham, Switzerland: Springer, 2022, pp. 696–706.

[8] G. Cicirelli, R. Marani, A. Petitti, A. Milella, and T. D'Orazio, "Ambient assisted living: A review of technologies, methodologies and future perspectives for healthy aging of population," *Sensors*, vol. 21, no. 10, p. 3549, May 2021.

[9] V. Vijayan, J. P. Connolly, J. Condell, N. McKelvey, and P. Gardiner, "Review of wearable devices and data collection considerations for connected health," *Sensors*, vol. 21, no. 16, p. 5589, Aug. 2021.

[10] E. Ramanujam, T. Perumal, and S. Padmavathi, "Human activity recognition with smartphone and wearable sensors using deep learning techniques: A review," *IEEE Sensors J.*, vol. 21, no. 12, pp. 13029–13040, Jun. 2021.

[11] A. Mallol-Ragolta, A. Semertzidou, M. Pateraki, and B. Schuller, "HarAGE: A novel multimodal smartwatch-based dataset for human activity recognition," in *Proc. 16th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Dec. 2021, pp. 1–7.

[12] S. Wang et al., "Technology to support aging in place: Older adults' perspectives," *Healthcare*, vol. 7, no. 2, p. 60, Apr. 2019.

[13] Z. Uddin, "Sensors and features for assisted living technologies," in *Applied Machine Learning for Assisted Living*. Cham, Switzerland: Springer, 2022, pp. 15–61.

[14] A. Raghav and S. Chaudhary, "Elderly patient fall detection using video surveillance," in *Proc. Int. Conf. Comput. Vis. Image Process.* Cham, Switzerland: Springer, 2022, pp. 450–459.

[15] M. Eldib, F. Deboeverie, W. Philips, and H. Aghajan, "Behavior analysis for elderly care using a network of low-resolution visual sensors," *J. Electron. Imag.*, vol. 25, no. 4, Mar. 2016, Art. no. 041003.

[16] C. Berridge and T. F. Wetle, "Why older adults and their children disagree about in-home surveillance technology, sensors, and tracking," *Gerontologist*, vol. 60, no. 5, pp. 926–934, Jul. 2020.

[17] P. Wang, W. Li, Z. Gao, J. Zhang, C. Tang, and P. O. Ogunbona, "Action recognition from depth maps using deep convolutional neural networks," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 4, pp. 498–509, Aug. 2016.

[18] J. Kim, K. Min, M. Jung, and S. Chi, "Occupant behavior monitoring and emergency event detection in single-person households using deep learning-based sound recognition," *Building Environ.*, vol. 181, Aug. 2020, Art. no. 107092.

[19] J. M. Sim, Y. Lee, and O. Kwon, "Acoustic sensor based recognition of human activity in everyday life for smart home services," *Int. J. Distrib. Sensor Netw.*, vol. 11, no. 9, Sep. 2015, Art. no. 679123.

[20] M. S. Khan, M. Yu, P. Feng, L. Wang, and J. Chambers, "An unsupervised acoustic fall detection system using source separation for sound interference suppression," *Signal Process.*, vol. 110, pp. 199–210, Jan. 2015.

[21] M. Schmitter-Edgecombe and D. J. Cook, "Assessing the quality of activities in a smart environment," *Methods Inf. Med.*, vol. 48, no. 5, pp. 480–485, 2009.

[22] D. J. Cook, A. S. Crandall, B. L. Thomas, and N. C. Krishnan, "CASAS: A smart home in a box," *Computer*, vol. 46, no. 7, pp. 62–69, 2012.

[23] H. Ghayvat et al., "Smart aging system: Uncovering the hidden wellness parameter for well-being monitoring and anomaly detection," *Sensors*, vol. 19, no. 4, p. 766, Feb. 2019.

[24] G. Santos, P. Endo, K. Monteiro, E. Rocha, I. Silva, and T. Lynn, "Accelerometer-based human fall detection using convolutional neural networks," *Sensors*, vol. 19, no. 7, p. 1644, Apr. 2019.

[25] E. Fridriksdottir and A. G. Bonomi, "Accelerometer-based human activity recognition for patient monitoring using a deep neural network," *Sensors*, vol. 20, no. 22, p. 6424, Nov. 2020.

[26] Y. Tang, Y. Tian, J. Lu, J. Feng, and J. Zhou, "Action recognition in RGB-D egocentric videos," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 3410–3414.

[27] S. Michibata, K. Inoue, M. Yoshioka, and A. Hashimoto, "Cooking activity recognition in egocentric videos with a hand mask image branch in the multi-stream CNN," in *Proc. 12th Workshop Multimedia Cooking Eating Activities*, Jun. 2020, pp. 1–6.

[28] G. M. Weiss, K. Yoneda, and T. Hayajneh, "Smartphone and smartwatch-based biometrics using activities of daily living," *IEEE Access*, vol. 7, pp. 133190–133202, 2019.

[29] S. Qiu et al., "Multi-sensor information fusion based on machine learning for real applications in human activity recognition: State-of-the-art and research challenges," *Inf. Fusion*, vol. 80, pp. 241–265, Apr. 2022.

[30] M. Sun et al., "EButton: A wearable computer for health monitoring and personal assistance," in *Proc. 51st ACM/EDAC/IEEE Design Autom. Conf. (DAC)*, Jun. 2014, pp. 1–6.

[31] J. Shi, D. Zuo, and Z. Zhang, "Transition activity recognition system based on standard deviation trend analysis," *Sensors*, vol. 20, no. 11, p. 3117, May 2020.

[32] J. Saha, C. Chowdhury, D. Ghosh, and S. Bandyopadhyay, "A detailed human activity transition recognition framework for grossly labeled data from smartphone accelerometer," *Multimedia Tools Appl.*, vol. 80, no. 7, pp. 9895–9916, Mar. 2021.

[33] S. A. Okour, "Classification of common basic activities of daily living using a rule-based system," Ph.D. dissertation, School Comput., Eng. Math., Western Sydney Univ., Penrith, NSW, Australia, 2015.

[34] M. Diyan, M. Khan, B. Nathali Silva, and K. Han, "Scheduling sensor duty cycling based on event detection using bi-directional long short-term memory and reinforcement learning," *Sensors*, vol. 20, no. 19, p. 5498, Sep. 2020.

[35] R. Possas, S. P. Caceres, and F. Ramos, "Egocentric activity recognition on a budget," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5967–5976.

[36] N. Starliper, F. Mohammadzadeh, T. Songkakul, M. Hernandez, A. Bozkurt, and E. Lobaton, "Activity-aware wearable system for power-efficient prediction of physiological responses," *Sensors*, vol. 19, no. 3, p. 441, Jan. 2019.

[37] F. Liang et al., "Collaborative fall detection using a wearable device and a companion robot," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2021, pp. 3684–3690.

[38] F. Erivaldo Fernandes, H. M. Do, K. Muniraju, W. Sheng, and A. J. Bishop, "Cognitive orientation assessment for older adults using social robots," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2017, pp. 196–201.

[39] *Telegram*. Accessed: Feb. 16, 2024. [Online]. Available: https://telegram.org

[40] A. Guenounou, M. Aillerie, A. Mahrane, M. Bouzaki, S. Boulouma, and J.-P. Charles, "Human home daily living activities recognition based on a LabVIEW implemented hidden Markov model," *Multimedia Tools Appl.*, vol. 80, pp. 24419–24435, Apr. 2021.

[41] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Jan. 1989.

[42] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1251–1258.

[43] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[44] B. Logan et al., "Mel frequency cepstral coefficients for music modeling," in *Proc. ISMIR*, vol. 270, no. 1. Plymouth, MA, USA, 2000, p. 11.

[45] Y. Chen and Y. Xue, "A deep learning approach to human activity recognition based on single accelerometer," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2015, pp. 1488–1492.

[46] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

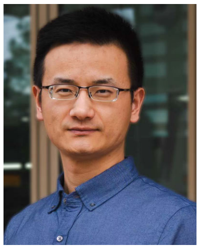
[47] *Socketio*. Accessed: Feb. 16, 2024. [Online]. Available: https://socket.io/docs/v4/

[48] *Casas Datasets*. Accessed: Feb. 16, 2024. [Online]. Available: http://casas.wsu.edu/datasets/

[49] J. Shahid, R. Ahmad, A. K. Kiani, T. Ahmad, S. Saeed, and A. M. Almuhaideb, "Data protection and privacy of the Internet of Healthcare Things (IoHTs)," *Appl. Sci.*, vol. 12, no. 4, p. 1927, Feb. 2022.

Fei Liang (Graduate Student Member, IEEE) received the B.S. degree in network engineering from Harbin University of Science and Technology, Harbin, China, in 2012, and the M.S. degree in computer science from Beijing University of Posts and Telecommunications, Beijing, China, in 2015. He is currently pursuing the Ph.D. degree in electrical and computer engineering with Oklahoma State University, Stillwater, OK, USA.

His research interests include wearable computing, machine learning, IoT, and robotics.



Zhidong Su (Member, IEEE) received the B.S. degree in mechanical engineering from Henan University of Technology, Zhengzhou, China, in 2016, and the M.S. degree in mechanical engineering from Guizhou University, Guiyang, China, in 2019. He is currently pursuing the Ph.D. degree in electrical and computer engineering with Oklahoma State University, Stillwater, OK, USA.

His research interests include human–robot interaction, reinforcement learning, natural language processing, computer vision, embedded systems, and deep learning.

Weihua Sheng (Senior Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Zhejiang University, Hangzhou, China, in 1994 and 1997, respectively, and the Ph.D. degree in electrical and computer engineering from Michigan State University, East Lansing, MI, USA, in May 2002.

He is currently a Professor at the School of Electrical and Computer Engineering, Oklahoma State University (OSU), Stillwater, OK, USA, where he is also the Director of the Laboratory for Advanced Sensing, Computation and Control (ASCC Lab, http://ascclab.org). He has authored more than 240 peer-reviewed papers in major journals and international conferences. Eight of them have received the best paper or best student paper awards in major international conferences. His current research interests include social robotics, wearable computing, human–robot interaction, and intelligent transportation systems. His research has been supported by the U.S. National Science Foundation (NSF), Department of Defense (DoD), and Oklahoma Transportation Center (OTC)/Department of Transportation (DoT).

Dr. Sheng served as an Associate Editor for IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING from 2013 to 2019. He is currently an Associate Editor of *IEEE Robotics and Automation Magazine*.