Collaborative Fall Detection using a Wearable Device and a Companion Robot

Fei Liang, Ricardo Hernandez, Jiaxing Lu, Brandon Ong, Matthew Jackson Moore, Weihua Sheng, Senlin Zhang

***Abstract*— Older adults who age in place face many health problems and need to be taken care of. Fall is a serious problem among elderly people. In this paper, we present the design and implementation of collaborative fall detection using a wearable device and a companion robot. First, we developed a wearable device by integrating a camera, an accelerometer and a microphone. Second, a companion robot communicates with the wearable device to conduct collaborative fall detection. The robot is also able to contact caregivers in case of emergency. The collaborative fall detection method consists of motion data based preliminary detection on the wearable device and video-based final detection on the companion robot. Both convolutional neural network (CNN) and long short-term memory (LSTM) are used for video-based fall detection. The experimental results show that the overall accuracy of video-based algorithm is 84%. We also investigated the relation between the accuracy and the number of image frames. Our method improves the accuracy of fall detection while maximizing the battery life of the wearable device. In addition, our method significantly increases the sensing range of the companion robot.**

I. MOTIVATION

In recent years, we have witnessed a steady growth of older adult population who are above the age of 65 [1]. This can be attributed to the fact that the baby boomer generation finally reached this age group while people's life expectancy has increased in recent decades. Older adults face many health problems and have to find assistance from the younger generation or professional caregivers. Particularly, fall has become a serious problem among older adults [2], which usually results in injuries, and sometimes even deaths. Due to financial reasons, it is not realistic to have a personal health practitioner to monitor the state of those facing health problems. Although many research efforts have been devoted to fall detection, accurately detecting a fall and providing medical care in real time is still very challenging.

This project is supported by the National Science Foundation (NSF) Grants CISE/IIS 1910993, EHR/DUE 1928711, the Joint Fund of Chinese Ministry of Education for Pre-research of Equipment under Grant 6141A02022371, and the Open Research Project of the State Key Laboratory of Industrial Control Technology, Zhejiang University, China (No. ICT2021B14).

Fei Liang, Ricardo Hernandez, Jiaxing Lu, Brandon T. Ong and Weihua Sheng are with the School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK, 74078, USA (e-mails: weihua.sheng@okstate.edu). Matthew Jackson Moore is with the School of Mechanical and Aerospace Engineering, Oklahoma State University, Stillwater, OK, 74078, USA. Senlin Zhang is with the State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou 310027, China. (e-mail: slzhang@zju.edu.cn).

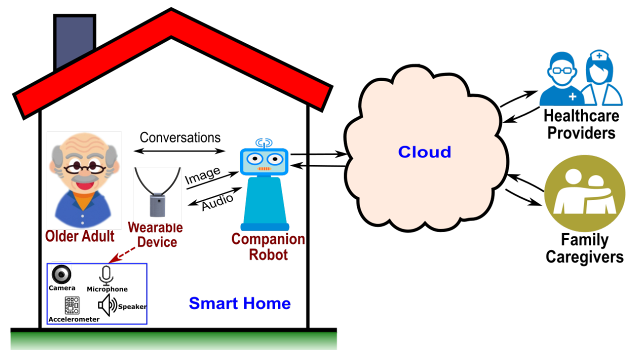

Fig. 1: The concept of collaborative activity monitoring using a wearable device and a companion robot.

In recent years, companion robots are coming to our homes which may provide a great opportunity to address the fall detection problem [3]. Companion robots can not only provide basic functions such as news, playing music, chatting, but also offer medical consultation and ask caregivers for help when older adults have an emergency. However, using cameras on the robot or in the home environment to monitor older adult activities is intrusive and may cause significant privacy concerns. Therefore we need to develop a new way for the robot to collect data to understand older adults behaviors and their surroundings. We propose that a wearable multi-sensor device could alleviate privacy concerns of users since the captured images are not of the user but of the surroundings [4]. In an independent living environment, this method can maximize privacy protection. In addition, when the wearable camera is integrated with other sensors such as a microphone and an accelerometer, the wearable device could collect more data that be useful for the robot to expand its sensing capability in understanding the behavior of the older adults.

Therefore, the objective of this paper is to propose, develop and test a collaborative activity monitoring system (CAMS) that combines a wearable device with a companion robot for elderly care. The current focus of activity monitoring is on fall detection. Fig. 1 illustrates the concept of collaborative activity monitoring in which the robot works closely with the wearable device to understand the behavior of the older adults. In emergency situations, such as a fall, the robot will notify the healthcare providers or family caregivers through a Cloud based management system.

This paper has three major contributions. First, it proposed and developed a collaborative strategy between a wearable device and a companion robot to understand human activities which minimizes the privacy intrusion while allowing the robot to maximize its sensing range. Second, a compact wearable activity monitoring unit (WAMU) is designed and implemented, which collects multi-modal data regarding the wearer and the environment, conducts preliminary fall detection based on motion data in a power-efficient way. Third, we developed, implemented and tested the vision-based fall detection algorithm by leveraging deep neural networks on the companion robot.

The rest of this paper is organized as follows: Section II introduces the related work. Section III presents the overall design of the CAMS. Section IV describes the design of the WAMU, the companion robot and the software. Section V presents the experimental setup, design and results. Section VI concludes our work and discusses the future work.

II. Related Work

A significant amount of work has been done recently in the field of fall detection using wearable devices. Several projects are based on a single sensor. Bourke *et al.* [5] used accelerometer to get data and develop a simple rule to classify an event as a fall by using a threshold which is derived by taking the magnitude of the three signals from each tri-axial accelerometer data. Torti *et al.* [6] focused on the implementation of recurrent neural networks (RNNs) architectures of microcontroller units (MCU) for fall detection with tri-axial accelerometers. Shojaei *et al.* [7] demonstrated a vision-based detection method using LSTM to verify a fall with 3D joint skeleton features. However, using a single sensor to collect data has limited information and usually results in many false positives.

Multiple wearable sensors were also used for fall detection, which usually offers higher accuracy. Hussain *et al.* [8] presented a system which includes a gyroscope and an accelerometer to minimize false positives in fall detection. Mao *et al.* [9] proposed a fall monitoring system in which a portable sensor collects the data with a 3-axis accelerometer, a 3-axis gyroscope and a 3-axis magnetometer then sends it to the mobile phone for further fall detection decisions. Ozcan *et al.* [4] developed a fall detection system using a wearable device equipped with cameras and accelerometers. They detected falls by analyzing the video and accelerometer data captured by a smartphone. As the smartphone needs to be attached to the waist with belt, it is not human-friendly for older adults. In addition, the camera and accelerometer are independent and can not communicate with each other, which may lead to unnecessary waste of energy due to the heavy computation on video data. Shahiduzzaman *et al.* [11] designed a Smart Helmet integrated with wearable cameras, accelerometers and gyroscope sensors, which is not convenient to the older adults living in home. They provided a fall detection algorithm that processes both accelerometer and video data on the Smart Helmet.

Other researchers explored the use of both wearable sensors and environmental sensors for fall detection. For example, Martinez *et al.* [12] proposed a multimodal fall detection system which consists of wearable sensors, ambient sensors and vision sensors. They combined LSTM with CNN to detect a fall from raw data (sensor data and video data) in real-time. However, fall detection based on environmental vision sensors causes significant privacy concerns.

Companion robots have been used to support elderly care in homes or care institutions [13]. Many companion robots have been developed for elderly care, including the Aibo [14] robot and Paro [15] robot. Equipped with various functions such as speech recognition, object recognition, and dialogue management, companion robots could be an avatar representing caregivers or social companions [3] [16] that improve the quality of life for older adults through companionship and social interaction. A companion robot can offer proactive assistance when emergency occurs. They could provide safety assistance including confirming a fall and sending notifications. Do *et al.* [3] developed a fall detection system based on sound events on a companion robot. It could recognize various falling sounds. After a fall is detected, the robot could connect to a remote caregiver for assistance. However, the shortcoming with the method is that the robot may not be able to hear the sound when the older adult is far from the robot.

III. Overall Design of Collaborative Activity Monitoring System (CAMS)

As shown in Fig. 2, the CAMS consists of a WAMU, a companion robot, and a healthcare management system.

The WAMU measures the acceleration of the user motion to determine if a fall is occurring. A positive result triggers the capture of a sequence of images that are then sent via WiFi to the companion robot. The companion robot processes the images to further confirm or reject the fall. If confirmed, the robot begins its emergency protocol, engaging the user to determine if the caregivers should be notified. The audio is recognized through the Google Speech API, while the Rasa [17] module provides basic skills for older adults to use, such as weather, news, quotes, jokes, music, photo-taking, and wiki information. The WAMU and the companion robot are also in constant communication so that the user can still take advantage of the capabilities of the robot.

The design requirements of the WAMU are as follows:

1) The WAMU should collect motion of the wearer and the video and audio data of the surrounding environment of the wearer.

2) The wearable device should be lightweight and ergonomic for older adults and rechargeable with a reasonable battery life for daily use.

3) The WAMU should be able to communicate with the companion robot within the range of a typical home, which allows the wearable device and the robot to collaboratively understand the user's daily activities.

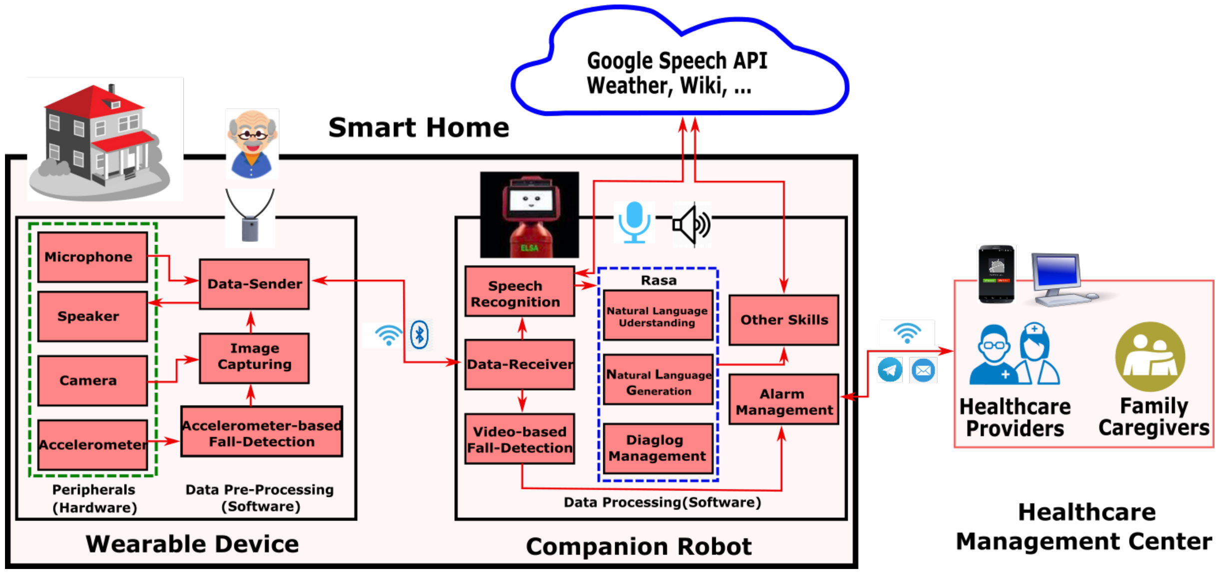

Fig. 2: The overall design of the CAMS.

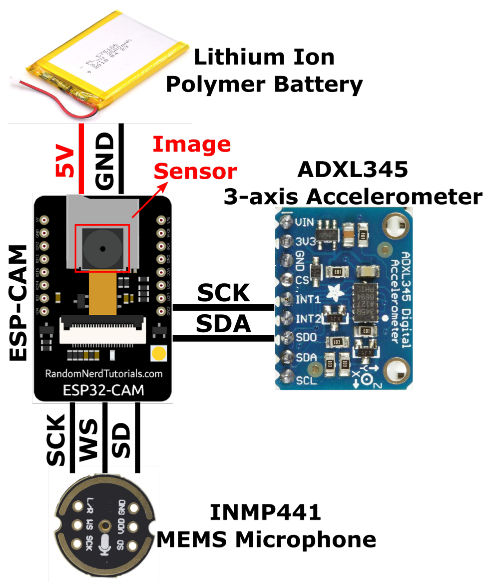

Fig. 3: The design of the Wearable Activity Monitoring Unit (WAMU).

IV. Hardware and Software Design

A. Wearable Activity Monitoring Unit

Fig. 3 shows the design of the WAMU which has three parts: the circuit board, battery and housing.

1) Circuit Board: As shown in Fig. 3, the board consists of an ESP32 development board kit, a 3-axis accelerometer, and a digital MEMS microphone. ESP32-CAM is a small embedded computing module that incorporates an ESP32S chip, an OV2640 CMOS image sensor, and an external 4MB PSRAM. This module is chosen for its compactness, the amount of available pins for use, and many other functionalities. Since the WAMU is worn by older adults for a sustained period of time, a lightweight design with a small form factor is desired. The general purpose IO pins allow multiple peripheral devices to be connected to the ESP32-CAM. The chip also integrates WiFi 802.11b/g/n allowing the board to connect to a server which runs on the companion robot. Finally with the extension of the memory capacity, large blocks of data such as images and audio can be acquired and stored on the board for preprocessing and transmission. The ADXL345 digital accelerometer is adopted as the motion sensor, which is a small, ultra-low power, 3-axis accelerometer with a high resolution of 13-bit measurements up to 16g. The sensitivity and resolution are sufficient for detecting falls. I2C communication protocol is used for the ADXL345 to talk to the ESP32-CAM at a rate of 800 Hz. Finally, the audio input device is an INMP441 omnidirectional MEMS microphone. This chip is fitted with the necessary features to provide a high-performance, low-power, digital-output using the industry-standard I2S protocol.

2) Battery: As shown in Fig. 3, the battery fitted to the circuit board is a lithium-ion polymer battery which provides a thin, lightweight design with a capacity of 2500 mAh. The output of the battery ranges from 4.7V at full charge to 3.7V. This voltage output is not compatible with the ESP32-CAM that requires a supply voltage of around 5V. To accommodate it, a voltage booster is used to increase the voltage to 5V. Along with the booster, the board is also equipped with a charging circuit that takes a USB-C connection and charges the battery at a max rate of 1000 mA.

It is necessary to have an estimate of the battery life. The component that has a considerable amount of power draw is the ESP32-CAM which has a maximum of 240 mA current consumption when transmitting data. The other components have very small power consumption. Therefore only the ESP32-CAM is considered in estimating the battery life. We assume 15 false positives will occur every hour. The average amount of current draw would be around 82.5 mA which results in a battery life of 30 hours and is sufficient

Fig. 4: The overview of the WAMU housing.

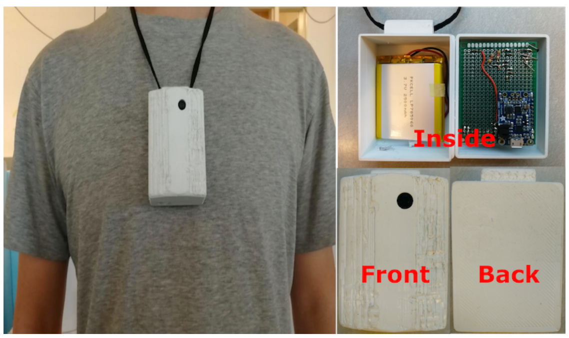

Fig. 5: A person wearing the wearable activity monitoring unit (WAMU).

for supporting daily use.

3) WAMU Housing: Many methods to house the wearable device hardware were investigated. Some of the designs include a solid frame necklace, a headset mount, and various corded necklace designs. After hands-on experimentation and rapid prototyping, the corded necklace proved to be the most adaptable and user friendly. The chosen design houses all electrical components in one case as shown in Fig. 4. The case has both front and back parts. On the housing there is a cut out for the camera lens (front) and charging port (bottom). During the prototyping phase, the housing can be opened and closed easily. A flexible cord is used to suspend the device around the user's neck. Compared to a solid frame, the cord allows the device to fit on any user. Fig.5 shows a person wearing the WAMU along with the inside, front and back view of the WAMU.

B. Companion Robot

Fig. 6 shows the ASCC Companion Robot [16] developed in our lab. It consists of three main parts: the head, the body and the power base. The robot head is comprised of a vision system with an Intel Realsense RGB-D camera, an auditory system with four microphones for speech recognition and sound localization, and a touch screen which is connected to an ARM-based board running Android OS and used for user interfacing. An Intel NUC with a Core i5 processor is hosted in the robot body, which facilitates speech recognition, video-based fall detection and other skills, as well as the capability

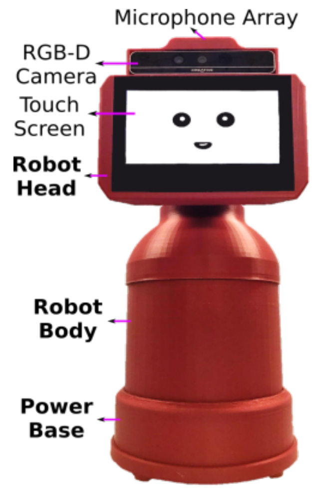

Fig. 6: The prototype of the ASCC companion robot.

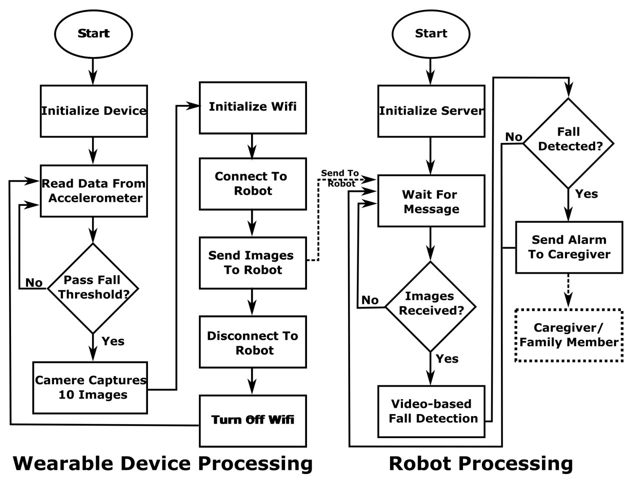

Fig. 7: The software flowchart of the proposed system.

to communicate with caregivers or family members when the older adult is in emergency situations.

C. Software

1) Software Flowchart: The flow chart of software is shown in Fig. 7. It can be divided into two parts: the wearable device part and the robot part. In the wearable device part, acceleration data are collected by the 3-axis accelerometer and the data are processed locally on the device. If the magnitude of the 3-axis acceleration vector is greater than the threshold, it is declared as a potential fall event and the camera function will be triggered. As a result, a sequence of images during falling are captured and sent to the robot. On the robot, after verifying the fall, an alarm will be sent to the caregivers or family members. Besides, the robot will communicate with the older adult through natural language. It will ask the older adult "are you ok?" and wait for older adults' responses to make further decisions.

2) Data Communication: The WAMU is connected to the robot using TCP/IP sockets. In order to save energy, the wifi turns on only when a possible fall event is detected. Then the image data are sent to the robot. After receiving the response from the robot, if it needs to respond to the robot by audio, the microphone data will be collected and sent through this connection. After that, the connection will be closed and the wifi will turn off.

3) Fall Detection Algorithms: Fall detection is implemented in two steps. The first step is a preliminary classifier that runs on the WAMU using the accelerometer data. The second step runs on the robot which receives a sequence of images from the WAMU and classifies them through recurrent neural network (RNN). The RNN-based classifier provides more accurate classification and reduces possible false positive results.

For the accelerometer-based preliminary fall detection a simple power-efficient algorithm is used, in which we compare the magnitude of the 3-axis acceleration vector against a predefined threshold for detecting a fall. According to [5], we set the threshold to 3g, which is sufficient to detect the falls. For the video-based algorithm, RNN is a type of artificial neural networks which consists of nodes from a directed graph along a temporal distance. Compared with convolutional neural network (CNN), the RNN model is influenced by previous features, which provides its memory of previous action and the ability to classify actions which involve consecutive image frames. We built an architecture for video classification based on both CNN and e long short-term memory (LSTM). We use the Resnet152 as the CNN for extracting features for each frame and compressing the frame from an image to a vector. The last fully-connected classification layer is deleted and therefore the result is a processed and compressed vector that contains the features of the original image and can be analyzed by the following RNN.

The LSTM is an artificial RNN architecture [18]. As shown in Fig. 8, compared with traditional RNN architectures, the LSTM architecture is more complicated. The RNN concatenates past state and current state and controls the outputs through the tanh function. The LSTM contains additional input and output. Thus, instead of only having memory of state at the most recent moment like the RNN, the LSTM has long time memory. After getting the processed feature vector through Resnet, an LSTM with three hidden layers analyses the vector and derives the classification result.

4) Alarm Management: As falls may result in serious injuries among the elderly [19], timely intervention by caregivers can play a vital role. We chose an open source mobile APP Telegram [20] to be the communication channel between the companion robot and the caregivers. With the mobile device, caregivers or family members can receive telegram messages when older adults are in trouble (e.g. fall or asking for help). The message type can be text, image or audio, which could provide abundant information for further decisions.

V. Experimental Evaluation

A. Test setup

We implemented the fall detection system using the proposed WAMU and the existing companion robot. The ESP32-

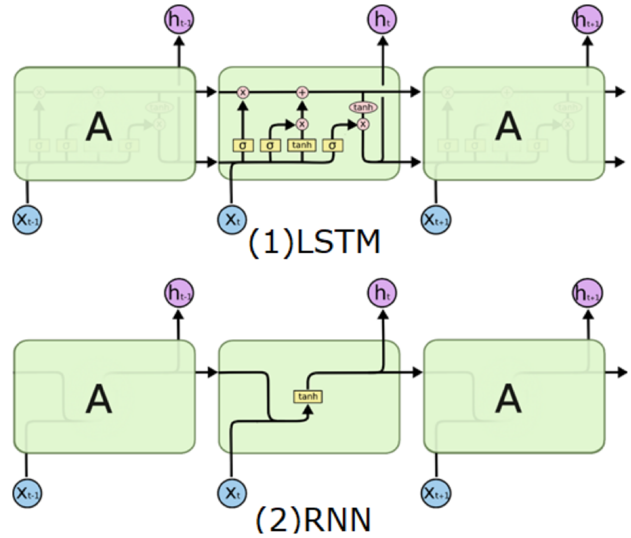

Fig. 8: The modules of LSTM and RNN.

Fig. 9: The smart home testbed and the testing scenario.

CAM board has a 2-core 160MHz CPU and 520KB SRAM plus 4MB PSRAM. The resolution of the captured images is 640 by 640 and the frame rate is between 15 and 60 fps. Meanwhile, the accelerometer-based fall detection algorithm runs on the board, and the accelerometer data has a sampling frequency of 400Hz. The robot runs on a 4-core Intel i5 CPU with 8G memory and the OS is Ubuntu 16.04. As shown in Fig. 5 and Fig. 9, in order to collect data and validate the algorithms, we set up a mattress in our smart home testbed. Human subjects participating in the experiment fell onto the mattress. Before the test, IRB approval was obtained from our university.

B. Experiment and Results

The wearable camera captures images and sends to the robot when it detects abnormal acceleration data. The fall detection program on the robot first compresses every image into a vector. The RNN then receives a sequence of input vectors representing each frame. A long short-term memory (LSTM) network is used to train the RNN model. In order to test the effectiveness of the RNN-based fall detection system,

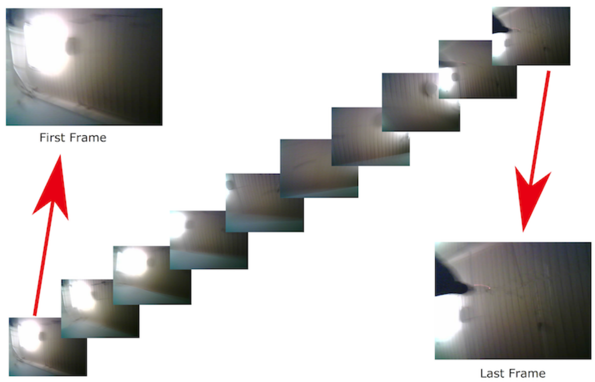

Fig. 10: A sample sequence of 10 consecutive images collected during a fall.

as Fig. 5 shows, we collected data from 4 male volunteers aged between 24 and 31. They were asked to complete each action in Table I. Each video represents an action and consists of 10 frames, as shown in Fig. 10. The wearable device captures the image of the environment instead of the user which helps reduce privacy concern. Six actions are listed for classification, as shown in Table I below. Hence, a total of 600 sets of experimental data were collected, with 100 sets for each action. We divided the dataset, which has a total of 6000 images, into three parts: training dataset, validation dataset, and test dataset, with the ratio of 3:1:1. In order to determine the suitable number of frames for fall detection, we compared the accuracy of fall detection using different numbers of frames.

TABLE I: The Dataset.

Action	Samples	Number of Frames
Walk	100	1000
Sit down	100	1000
Fall forward	100	1000
Fall backward	100	1000
Fall left	100	1000
Fall right	100	1000
Total	600	6000

TABLE II: Confusion Matrix of Detection using Resnet + RNN

Predicted Classes \ Real Classes	Backward	Forward	Left	Right	Sit down	Walk
Backward	0.9434	0.0138	0.0446	0.0235	0.0290	0.0032
Forward	0.0081	0.8489	0.0250	0.0451	0.0159	0.0021
Left	0.0232	0.0277	0.7902	0.1294	0.0654	0.0284
Right	0.0182	0.0723	0.1027	0.7373	0.0813	0.0221
Sit down	0.0020	0.0255	0.0232	0.0147	0.7935	0.000
Walk	0.0051	0.0117	0.0143	0.0500	0.0150	0.9442

The confusion matrix of our proposed approach on testing dataset is shown in Table II. The overall accuracy is 84%, as Fig. 11 shows, and it is obtained at 53th epoch. As shown in the table, the accuracy of some classes reaches 94%, including falling backward and walking. However, the accuracy of falling left, falling right, and sitting down is not as good as that of other classes.

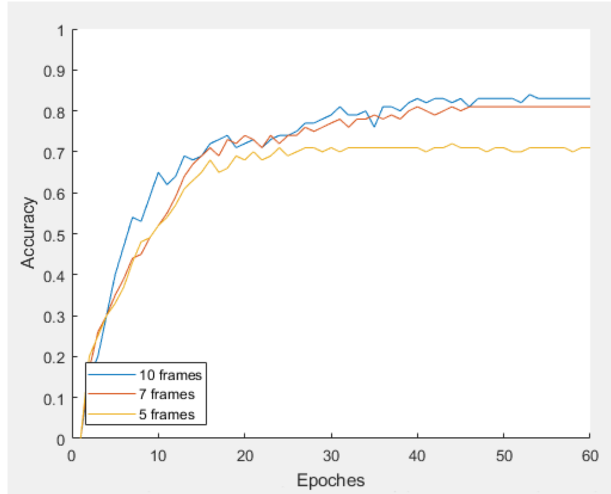

Fig. 11: Accuracy of Resnet+RNN with respect to the number of image frames used.

From Fig. 11, we find that using 10 frames the accuracy of fall detection is 84%. When the frame number is less than 5, the accuracy is too low to verify a fall. The accuracy of using 7 frames is very close to the accuracy of using 10 frames. But with 7 frames, we can save 3 frames which leads to lower power consumption in communication and computation.

VI. Conclusion and Future Work

In this paper, we presented the design and implementation of a collaborative activity monitoring system (CAMS) based on a wearable device and a companion robot. We use fall detection as a case study to evaluate the CAMS. The collaborative fall detection method consists of two steps: accelerometer-based fall detection on the WAMU and video-based fall confirmation or rejection on the companion robot. Accelerometer-based detection algorithm is energy-efficient, which extends the battery life of the wearable device. The video-based algorithm combines CNN and LSTM. The experimental results show that the overall accuracy of fall detection using the video-based algorithm is 84%, while for some classes of activities the accuracy reaches 94%, including falling backward and walking. We also analyzed the accuracy of fall detection using different numbers of image frames, which achieves a tradeoff between accuracy and power consumption. Apparently, the results we reported here are still preliminary but promising. We need further improve the whole system, particularly in the following aspects: 1) Energy consumption. We will conduct a more thorough investigation of the power consumption on the WAMU and optimize both hardware and software to further reduce the power consumption. 2) Fall detection. We will collect more data from more realistic falls and study the impact of various factors on fall detection accuracy, including the speed of falling, the location of falling, etc. 3) WAMU design. We will further improve the ergonomics of the WAMU to make it more human-friendly.

References

[1] J. Zhao and X. Li, "The status quo of and development strategies for healthcare towns against the background of aging population," *Journal of Landscape Research*, vol. 10, no. 4, pp. 41–44, 2018.

[2] T. Nguyen Gia, V. K. Sarker, I. Tcarenko, A. M. Rahmani, T. Westerlund, P. Liljeberg, and H. Tenhunen, "Energy efficient wearable sensor node for IoT-based fall detection systems," *Microprocessors and Microsystems*, vol. 56, no. October 2017, pp. 34–46, 2018. [Online]. Available: https://doi.org/10.1016/j.micpro.2017.10.014

[3] H. M. Do, M. Pham, W. Sheng, D. Yang, and M. Liu, "RiSH: A robot-integrated smart home for elderly care," *Robotics and Autonomous Systems*, vol. 101, pp. 74–92, 2018.

[4] K. Ozcan and S. Velipasalar, "Wearable Camera- and Accelerometer-Based Fall Detection on Portable Devices," *IEEE Embedded Systems Letters*, vol. 8, no. 1, pp. 6–9, 2016.

[5] A. K. Bourke, J. V. O'Brien, and G. M. Lyons, "Evaluation of a threshold-based tri-axial accelerometer fall detection algorithm," *Gait and Posture*, vol. 26, no. 2, pp. 194–199, 2007.

[6] E. Torti, A. Fontanella, M. Musci, N. Blago, D. Pau, F. Leporati, and M. Piastra, "Embedded real-time fall detection with deep learning on wearable devices," *Proceedings - 21st Euromicro Conference on Digital System Design, DSD 2018*, pp. 405–412, 2018.

[7] A. Shojaei-Hashemi, P. Nasiopoulos, J. J. Little, and M. T. Pourazad, "Video-based Human TelegramFall Detection in Smart Homes Using Deep Learning," *Proceedings - IEEE International Symposium on Circuits and Systems*, vol. 2018-May, pp. 0–4, 2018.

[8] F. Hussain, F. Hussain, M. Ehatisham-Ul-Haq, and M. A. Azam, "Activity-Aware Fall Detection and Recognition Based on Wearable Sensors," *IEEE Sensors Journal*, vol. 19, no. 12, pp. 4528–4536, 2019.

[9] A. Mao, X. Ma, Y. He, and J. Luo, "Highly portable, sensor-based system for human fall monitoring," *Sensors (Switzerland)*, vol. 17, no. 9, 2017.

[10] A. Pinto, "Wireless Embedded Smart Cameras: Performance Analysis and their Application to Fall Detection for Eldercare," p. 122, 2011.

[11] K. M. Shahiduzzaman, X. Hei, C. Guo, and W. Cheng, "Enhancing Fall Detection for Elderly with Smart Helmet in a Cloud-Network-Edge Architecture," *2019 IEEE International Conference on Consumer Electronics - Taiwan, ICCE-TW 2019*, pp. 2019–2020, 2019.

[12] L. Martinez-VillasenoTelegramr, H. Ponce, and K. Perez-Daniel, "Deep learning for multimodal fall detection," *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, vol. 2019-October, pp. 3422–3429, 2019.

[13] I. Anghel, T. Cioara, D. Moldovan, M. Antal, C. D. Pop, I. Salomie, C. B. Pop, and V. R. Chifu, "Smart environments and social robots for age-friendly integrated care services," *International Journal of Environmental Research and Public Health*, vol. 17, no. 11, 2020.

[14] Aibo. [Online]. Available: http://www.sony-aibo.com/

[15] Paro. [Online]. Available: http://www.parorobots.com/

[16] F. Erivaldo Fernandes, H. M. Do, K. Muniraju, W. Sheng, and A. J. Bishop, "Cognitive orientation assessment for older adults using social robots," in *2017 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2017, pp. 196–201.

[17] Rasa. [Online]. Available: https://rasa.com/

[18] Lstm. [Online]. Available: http://colah.github.io/posts/2015-08-Understanding-LSTMs/

[19] M. Alwan, P. J. Rajendran, S. Kelli, D. Mack, S. Dalali, and M. W. I, "KellI, DalalI, FelderI," pp. 6–10, 2006.

[20] Telegram. [Online]. Available: https://telegram.org/