IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 28, 2021, accepted May 25, 2021, date of publication May 31, 2021, date of current version June 15, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3084926

A Novel Hybrid Deep Learning Model for Activity Detection Using Wide-Angle Low-Resolution Infrared Array Sensor

MUTHUKUMAR K. A.[1], **(Graduate Student Member, IEEE),**
MONDHER BOUAZIZI[2], **(Member, IEEE), AND**
TOMOAKI OHTSUKI[2], **(Senior Member, IEEE)**
[1]Ohtsuki Laboratory, Center for Information and Computer Science, School of Science for Open and Environmental Systems, Graduate School of Science and Technology Yagami Campus, Keio University, Yokohama 223-8522, Japan
[2]Faculty of Science and Technology, Keio University, Yokohama 223-8522, Japan

Corresponding author: Muthukumar K. A. (kumar@ohtsuki.ics.keio.ac.jp)

This work was supported by the Grant Keio Leading-Edge Laboratory of Science and Technology, Japan, under Grant KEIO-KLL-000079.

ABSTRACT In this paper, we propose a deep learning-based technique for activity detection that uses wide-angle low-resolution infrared (IR) array sensors. Alongside with the main challenge which is how to further improve the performance of IR array sensor-based methods for activity detection, throughout this work, we address the following challenges: we employ a wide-angle infrared array sensor with peripheral vision in comparison to a standard IR array sensor. This makes activities at different positions have different patterns of temperature distribution, making it challenging to learn these different patterns. In addition, unlike previous works, our goal is to perform the activity detection using the least possible amount of information. While the conventional works use a time window equal to 10 seconds where a single event occurs, we aim to identify the activity using a time window of less than 1 second. Nevertheless, we aim to improve over the accuracy obtained in previous work by employing deep learning, while keeping the approach light for it to run on devices with low computational power. Therefore, we use a hybrid deep learning model well suited for the classification of distorted images because the neural network learns the features automatically. In our work, we use two IR sensors (32 × 24 pixels), one placed on the wall and one on the ceiling. We collect data simultaneously from both the IR sensors and apply hybrid deep learning classification techniques to classify various activities including ''walking'', ''standing'', ''sitting'', ''lying'', ''falling'', and the transition between the activities which is referred to as ''action change''. This is done in two steps. In the first step, we classify ceiling data and wall data separately as well as the combination of both (ceiling and wall) using a Convolutional Neural Network (CNN). In the second step, the output of the CNN is fed to a Long Short Term Memory (LSTM) with a window size equal to 5 frames to classify the sequence of activities. Through experiments, we show that the classification accuracy of the ceiling data, wall data, and combined data with the LSTM reach 0.96, 0.95, and 0.97, respectively.

INDEX TERMS Activity detection, hybrid deep learning, AI healthcare, infrared array sensor.

I. INTRODUCTION

Population ageing is a societal issue facing many countries nowadays that affects not only social life but also the economy. As a matter of fact, advancements in healthcare and medicine have continuously increased the average life expectancy over the last few decades. Today, the total world population stands at 7.9 billion [1] with 703 million people above the age of 65. Asia and Europe account for most of the elderly population in the world. Japan, for instance, is at the very top, with 28% [2] of its population above the age of 65. This high ratio of elderly people and increase in life expectancy combined with the fact that most of these people are living alone have made it necessary to develop more sophisticated techniques and technologies to monitor them. In this regard, artificial intelligence (AI) [3] plays an

The associate editor coordinating the review of this manuscript and approving it for publication was Wenming Cao.

important role in healthcare, particularly in assistive care technologies [4] for old people owing to the spurt in the Internet of Things (IoT)-based technological applications. In assistive care technology, activity detection is one of the key tasks to prevent accidents that might occur to elderly people.

Activity detection is mostly based on devices that could broadly be categorized into two categories: wearable devices and non-wearable ones. Wearable devices require the person to wear them and carry them all times to monitor the activity effectively and accurately. Such devices include smartphones [5], smart watches [6], [7], etc. that make use of accelerometers [8], [9], kinetic sensors [10], etc. Wearing these devices all the time makes an uncomfortable experience for the elderly. There is also a risk of damage to the device if they fall accidentally. In such situations, non-wearable devices provide several advantages compared with wearable ones. For instance, they allow for avoiding any physical contact with the person, reducing the burden for the elderly. The non-wearable devices are based on technologies such as cameras, sensors, antennas [11], radars [12], etc., and are placed in specific locations to monitor the elderly. However, non-wearable devices have some serious disadvantages like privacy issues, issues related to coverage, etc. That being said, they are less invasive and less burdensome than wearable ones.

Non-wearable devices are more convenient for elderly people inspite of limitations such as privacy issues arising from the use of cameras, coverage issues arising from the use of radars, and compatibility issues among wireless sensors. These issues have largely been overcome with the recent introduction of the infrared array sensor. These sensors are less invasive and more convenient to use in indoor environments. Recently, the infrared array sensor [13] became popular in healthcare technologies. The infrared array sensor measures the heat generated from the human body and projects it on a low-resolution matrix which could then be visualized as an image. It has several advantages: non-invasive from a privacy perspective, ease of positioning/set-up, better coverage resulting in a wider area of detection, etc. Moreover, it's low cost makes it affordable to implement. These advantages make infrared array sensors economical for use in a variety of industries such as aerospace, healthcare, automotive, etc.

Many infrared array sensor-based activity detection have been proposed in the past few decades [13], [14] [15], [16] [17]. However, each of these activity detection systems has its own limitations. Most activity detection systems are based on conventional machine learning methods such as Support Vector Machine (SVM), k-Nearest Neighbors (k-NN), etc. These conventional methods extract activity features manually to identify activities. As a result, the identification of activity with different people is less accurate. Mashiyama *et al.* [13] proposed an activity and fall detection technique [14] with an infrared array sensor (8 × 8 pixels) mounted at the ceiling using the SVM and k-NN classifiers. This approach does not perform well on detecting certain activities, such as sitting, etc. Kobayashi *et al.* [15] proposed an activity detection system with two infrared array sensors, one on the ceiling and the other on the wall and classified the activities using the conventional machine learning method SVM. This approach was intended to improve on the previous one [13] by integrating the data obtained from both sensors. They achieved over 90% in the detection of all the activities. In particular, the detection of sitting activity increased from 78% to 93%. However, the detection of some other activities, including walking and falling, underperformed. Recently, Fan *et al.* [16] and Taniguchi *et al.* [17] proposed to detect activities using infrared array sensors placed on the ceiling and on the wall, respectively. Classifying activities such as walking, sitting, standing, etc. using Recurrent Neural Network (RNN) models achieved 85% and 93% accuracy, respectively. Based on these findings and the limitations of the previous systems, we strongly believe that the activity detection could be further improved, and more accurate systems could be built.

In this paper, we propose a hybrid deep learning technique to detect the activities using a wide-angle low-resolution infrared array sensors. The wide-angle sensor produces an image distortion, in that when the person is in the periphery of the sensor's field of view, the image of the subject captured is different from when the subject is closer to the sensor. From the distorted image, it is hard to extract features useful for classification using conventional machine learning methods. Therefore, we use a hybrid deep-learning model to classify the distorted image; the neural network automatically learns distorted images' features. Two sensors are used in the proposed system. One of the sensors is placed on the wall, and another one is placed on the ceiling. Both the sensors collect the data at eight frames per second and start simultaneously. After collecting the data, the proposed activity detection technique involves two stages. First, we classify the individual frames collected by the wall sensor and the ceiling sensor separately using a Convolutional Neural Network (CNN). In the second stage, the output of the CNN is passed through a Long Short Term Memory (LSTM) with a window size equal to 5 frames to classify the sequence of activities. Afterwards, we combine the ceiling data and wall data and classify each pair of frames using CNN. The output of the CNN is passed through the LSTM with a window size equal to 5 frames. This leads to an improvement of the classification accuracy of various activities thanks to combining both sensor data. The contributions of this paper follows:

1) We proposed an activity detection system using a hybrid deep learning model, which could classify the distorted image produce by the wide-angle IR arrays sensor regardless of the amount of distortion.
2) Participants performed all the activities in all possible positions within the sensor coverage area irrespective of the sensor position. Most of the existing work perform activity only in front of or under the sensor.

3) We identify the activity using a time window of less than 1 second. Despite the small time window, we have remarkably enhanced the classification accuracy compared to the conventional works which require a larger time window.
4) Our customized lightweight neural network can run on devices with low computation power.

The remainder of this paper goes as follows: In section II, we introduce some of the existing work related to activity detection using wearable and non-wearable devices. In section III, we describe our motivations for this work and some of the challenges we faced during this research. In section IV, we introduce our proposed framework, as well as the experiment specifications. In section V, we describe in details of our model architectures and classification. In section VI, we show results and performance evaluation. In section VII, we discuss the research findings and the future direction of the research. Finally, in section VIII, we conclude this work.

II. RELATED WORK

A. WEARABLE DEVICE

Anguita *et al.* [19] proposed an approach that uses smartphones for activity detection. In this approach, subjects must always hold the smartphone. Human activity signals are obtained from smartphone inertial sensors. Features are extracted manually from the signals received by the sensors and classified using SVM with an overall accuracy of 87%. Mannini and Sabatini [20] proposed a wearable accelerometer sensor-based approach to classify various physical activities. Features are derived from the data acquired by the accelerometer sensor based on the linear acceleration component due to body motion and gravitational acceleration component. After extracting the features, the classification is performed using various probabilistic and geometric approaches. The best performance was obtained by Hidden Markov Model (HMM)-based classifier. Balli *et al.* [7] proposed a smartwatch-based approach for human activity detection. The device must be worn by the subject and data are obtained from the sensor. Using various machine learning methods such as SVM, k-NN, and random forest algorithm, features are extracted from the collected data. This study shows that the Random Forest algorithm performs better than SVM and k-NN.

Wearable device-based methods have their performance varying quite widely based on the type of sensor and the machine learning algorithm used. Nevertheless, they have their own limitations. A common shortcoming among them, however, is the need for manual extraction of features. Furthermore, the inconvenience of carrying such devices continuously is a drawback inherent to wearable devices, and cannot be avoided.

B. NON-WEARABLE DEVICE

Mashiyama *et al.* [13] presented a fall and activity detection system [14] using an infrared array sensor (8 × 8 pixels) placed on the ceiling. Data is obtained from the sensor sequentially in a fixed time window size. Features used to classify the action are manually extracted. In total, four features are derived from the temperature distribution of the data, i.e., the consecutive frames where the motion is observed, the maximum number of pixels that changed during these consecutive frames, the maximum temperature pixel variance, and the maximum temperature pixel distance before and after the activity. The classification accuracy obtained by using the k-NN algorithm for fall detection is 94%, and using SVM is 100% for no activity, 94.8% for stopping, 99% for walking, and 78% for sitting. These results show clearly that activities such as sitting require further improvement. This method ignores the influence of the detection angle. In addition, this method classifies fixed scenarios and does not detect the transition between the activities.

Kobayashi *et al.* [15]. proposed an activity detection system using two infrared array sensors, one on the wall and one on the ceiling. SVM is used for the classification of various activities. The combination of two sensor data increases the accuracy of the sitting activity detection from 78% to 93% compared to the conventional method [14]. The average classification accuracy of this system is above 90% for all activities performed. This approach has a few limitations. To begin with, their method for feature extraction is less effective than the conventional ones and does not detect the transition between the activities.

Fan *et al.* [16] has proposed a robust fall detection system using an infrared array sensor (8 × 8 pixels). The sensor is placed on the wall in this system. The different activities are carried out in parallel and perpendicular to the sensor. The data is pre-processed by applying a Gaussian filter, and a median filter then forwarded to an LSTM and a Gated Recurrent Units (GRU) recurrent neural networks to be classified. The system achieved an accuracy equal to 75% using LSTM and 85% using GRU. The activities in this system are performed in limited positions, and the accuracy of the classification is low in both the algorithms.

Taniguchi *et al.* [17] has proposed a fall detection system using two thermal array sensors (16 × 16 pixels). One is placed on the wall, and the other on the ceiling. All the activities are carried out under and in front of the sensors. Both the sensor data are combined, and the temperature distribution is used to distinguish fall activities from non-fall activities. Their approach achieved an accuracy equal to 72%. This system, however, relies on one of the oldest time series analysis approaches like time-series posture transition diagram, and the sum of temperature distribution. Several newer machine learning models perform much better.

Taramasco *et al.* [18] has proposed a fall detection system using an infrared array sensor (1 × 16 pixels). The sensor is placed on the ceiling, and the subjects have carried out a variety of activities, the data of which are collected using the sensor. Activities are classified using a Recurrent Neural Network (RNN), which is used for classifying sequences with different architectures, such as LSTM, GRU, and Bi-LSTM

TABLE 1. Comparison of existing methods using infrared array sensor.

Study	IR sensor (resolution)	No. of sensors	Position of sensor	Methods	Accuracy	Limitations
Mashiyama et al. [13]	8 × 8	1	Ceiling	SVM	above 94%	Few activities in a specific area, no detection of transition between activities.
Mashiyama et al. [14]	8 × 8	1	Ceiling	k-NN	94%	Less effective feature extraction method.
Kobayashi et al [15]	8 × 8	2	Ceiling, Wall	SVM	above 90%	Activities are performed in very specific positions, difficulty to differentiate activities due to reactive pixels.
Xiyui et al. [16]	8 × 8	1	Wall	LSTM, GRU	75% and 85%	Very limited positions: activities are performed only in parallel or perpendicular to the sensor.
Taniguchi et al. [17]	16 × 16	2	Ceiling, Wall	Time series analysis	72%	Low accuracy due to the use of an old approach.
Taramasco et al. [18]	1 × 16	2	Opposite corner of the room	LSTM, GRU, Bi-LSTM	93%	High computation cost.

(Bi-directional LSTM). Their performance has varied widely. However, Bi-LSTM performed the best, achieving an accuracy equal to 93%. The Bi-LSTM approach requires a high computation device to run, limiting its usability on low computation devices.

Table 1 shows a comparison of the existing work and their limitations. Most of the existing work performs the activity in front of the sensor. They did not do the activities in different positions within the coverage areas of the sensor. The different angles in addition to the distortion of the images, produce different patterns for the same activity, making it hard for these activity detection systems to detect the activity, affecting their performance.

III. MOTIVATIONS AND CHALLENGES

There are a lot of non-wearable devices available for activity detection. They include but are not limited to radars, WiFi, infrared sensors, etc. Among these, we specifically choose to use the infrared array sensor because it has several advantages. Not only does it protect the privacy of the user, but it also operates in a variety of environments (in terms of luminance, including darkness). Most of the applications and existing work relying on infrared array sensors use ones with a resolution equal to 8×8 pixels. However, these sensors have very limited coverage and can be used only in a very small room.

Given the limitations, we listed above, we have chosen to use a 32 × 24 pixels resolution infrared array sensor. The sensor's resolution is higher than that of the existing 8 × 8, 1 × 16, 16×16 pixels resolution sensors. Nonetheless, the sensor that we are using has a wide-angle allowing for a coverage area that is much wider than that of the other sensors. The most challenging aspect of this activity detection using a wide-angle infrared array sensor is the distortion due to the wide-angle lens: for the same activity performed, when a person is distant from the sensor, his reflection on the captured image is much different from that when he is closer to the sensor. This makes this type of images non-appropriate for feature extraction using conventional machine learning methods. The features in these methods are extracted using temperature distribution changes. Clear images are easily classified using these conventional methods, whereas, distorted images are much harder to classify. This makes deep learning techniques more suitable for images of this kind. In the field of deep learning techniques, the activities' features are automatically learned by the neural network.

IV. EXPERIMENT SPECIFICATIONS

A. DEVICE SPECIFICATION

We used two of the MLX90640 (Melexis corporation)[1] infrared array sensor shown in FIGURE 1. These sensors are capable of detecting heat rays from any thermal source. Table 2 displays the main sensor specifications. The sensor temperature range covers both the typical human temperature as well as indoor temperature. Nevertheless, the sensor can collect data at different frame rates. The sensor frame resolution is 32×24 pixels. The brighter the color is in the generated frames, the higher the temperature is.

The sensor is attached to a Raspberry Pi 3 model b+ as shown in FIGURE 2. The Raspberry Pi is also equipped with a standard camera recording the same event as the sensor. The data collected by the camera are used as ground truth and are used to annotate the sensor data. We prepared two sets of devices with the same configuration, one is placed on the wall, and the other is placed on the ceiling. The wall sensor and the ceiling sensor as well as their corresponding cameras collect data at the same rate of 8 frames per second (fps). The data are stored in the SD card mounted in the Raspberry Pi.

[1] https://www.melexis.com/en/product/MLX90640/

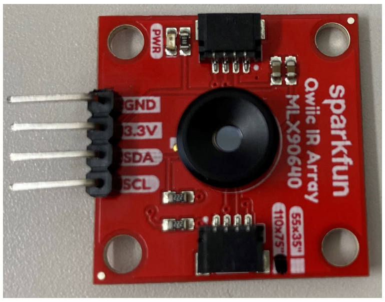


FIGURE 1. The wide angle infrared array sensor used for our experiments.

TABLE 2. The technical specifications of the sensor.

Infrared sensor model: Qwiic IR Array	MLX90640
Camera	1
Voltage	3.3 V
Temperature range of targets	$-40°C \sim 85°C$
Absolute temperature accuracy	$\pm 2°C$
Number of pixels	768 (32×24)
Viewing angle	110° × 75°
Frame rate	8 frames/second

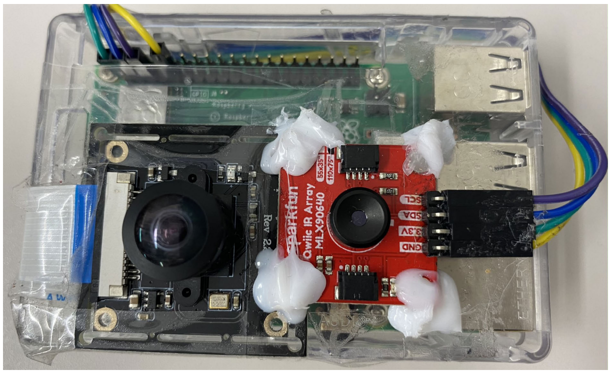


FIGURE 2. An image of the Rasberry Pi 3+ with the camera and the IR senors mounted which we used for collecting the data.

B. ENVIRONMENT

The experiment has been set up in a large meeting room environment with a standard room temperature. Two IR array sensors are deployed in the room, one on the ceiling and the other on the wall. In FIGURE 3, we show a simplified scheme of the sensor deployment and an example of a frame collected by the sensor.

FIGURE 4 shows the coverage measurements according to the sensor specification. The sensor has a wide-angle: the coverage alongside the first angle is 110°, and alongside the other is 75°. Using these angles, we calculate the ceiling sensor coverage area, i.e., length × breadth, which we refer to as l_1 and l_2, respectively (which correspond to the coverage and the angles θ_1 and θ_2, respectively). The sensor is attached to the ceiling at a height equal to 2.60 m from the ground. We refer to this height as h_c. Based on the known values, the coverage area can be calculated using the following equations.

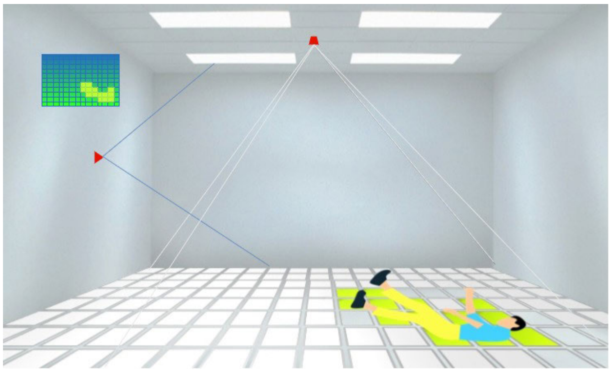


FIGURE 3. The actual coverage area of the sensor.

$$l_1 = 2 \cdot h_c \cdot \tan\left(\frac{\theta_1}{2}\right) \tag{1}$$

$$l_2 = 2 \cdot h_c \cdot \tan\left(\frac{\theta_2}{2}\right) \tag{2}$$

The coverage at the ground level, however, is not realistic. In addition, in the case where the human is at the edge, barely his feet will be detected, as shown in FIGURE 4. Therefore, we use α and β coefficients to ensure that coverage is sufficiently reliable. In consideration of our early experiments, α is set to be 0.81, and β is set to be 0.75. This will effectively cover an area whose length and breadth are equal to 7.40 m and 3.90 m, respectively.

The wall sensor is placed on the wall at a height 1.00 m from the floor, the height of the sensor represented in h_w. The coverage area of the wall sensor is shown in FIGURE 4. We calculated the wall sensor coverage area using the angle of the sensor θ_1 and the h_w. Here, γ is the angle between the sensor coverage and the wall. The blind angle of the sensor, where the detection using the wall sensor is not possible, is represented by the distance d. Based on the known values of θ_1 and h_w, we calculated d using the following equations.

$$\gamma = 90^\circ - \left(\frac{\theta_1}{2}\right) \tag{3}$$

$$d = h_{\mathrm{w}} \cdot \tan\gamma \tag{4}$$

We experimented in 2 different rooms. The first room is a small closed space, with a little amount of sunlight entering from its single window. The Air Conditioner (AC) temperature in the room is set to 24° C. The second room is wider, has a large window allowing more sunlight to enter the room, and has an AC whose temperature is set to 22° C. Five different people, males and females of different ages, participated in the experiments. In each experiment, a single person is asked to perform various activities contentiously for 5 minutes. Data are collected by the sensors, which we use later on for classification. We conducted several experiments and collected enough data for training and evaluating the proposed approach.

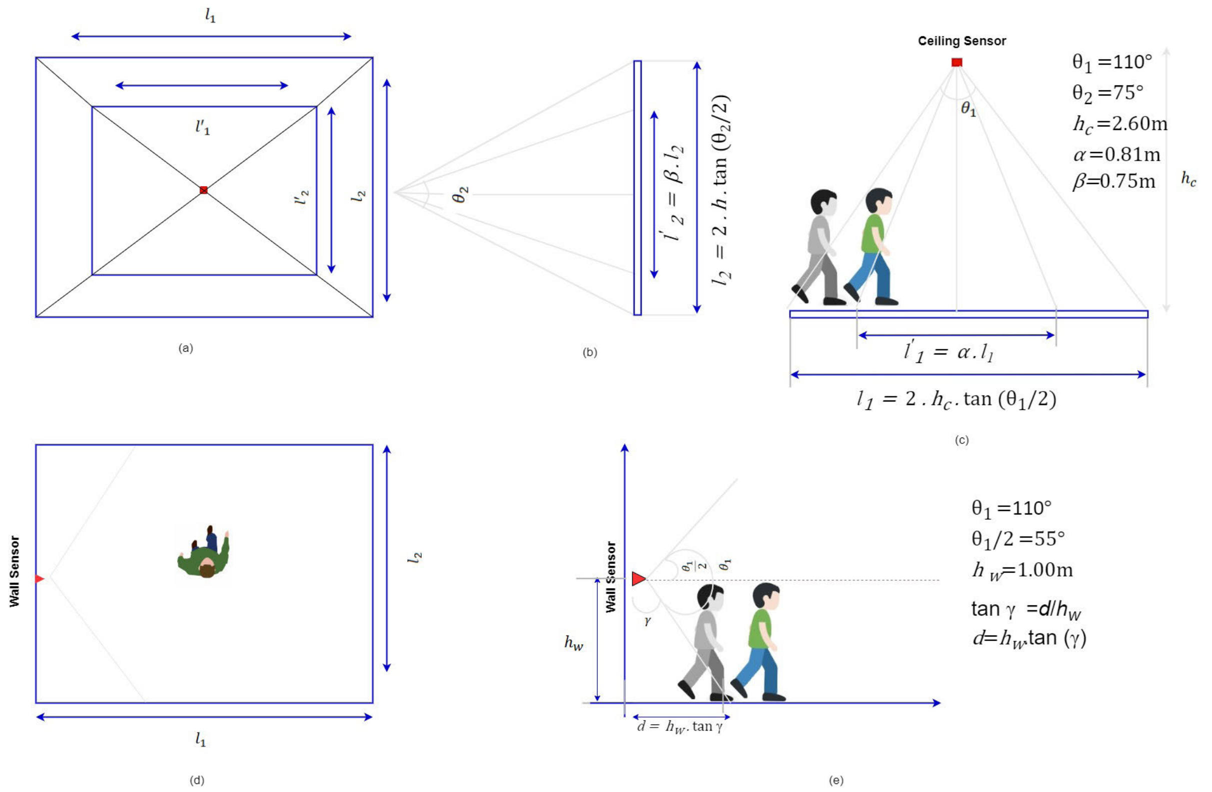


FIGURE 4. The area covered by the sensor and its detailed dimensions: (a) Top View of the ceiling sensor; (b) Side view of the ceiling sensor; (c) Front view of the ceiling sensor and its calculated dimensions; (d) Top view of the wall sensor; (e) Side view of the wall sensor and its calculated dimensions.

C. FRAMEWORK

The overall framework of our proposed system is shown in FIGURE 5. We collect data from our experiments, and then using CNN and LSTM, we perform the classification of the various activities. First, we classify individual frames collected by the wall sensor and the ceiling sensor separately using the CNN. We then pass the output of the CNN through the LSTM for sequential activity classification and check the performance on both the ceiling sensor data and the wall sensor data separately. Second, we combine the wall sensor data and the ceiling sensor data. Using CNN, we classify the individual pairs of frames of the activities and analyze the performance. The output of the CNN is passed through the LSTM for sequential classification of the activities. The outputs of CNN and LSTM using wall sensor data are represented by $\mathrm{CNN_w}$ and $\mathrm{LSTM_w}$, respectively. In the same way, the outputs of CNN and LSTM using ceiling data are represented by $\mathrm{CNN_c}$ and $\mathrm{LSTM_c}$, respectively. The output of CNN and LSTM using the combined ceiling sensor data and wall sensor data is represented by $\mathrm{CNN_{cw}}$ and $\mathrm{LSTM_{cw}}$, respectively. We use these notations to differentiate between the different models and to make it easy to compare their performance. Some samples of activities as well as their predictions using the CNN and the CNN+LSTM networks are given in FIGURE 6. In FIGURE 6(a), we show examples of frames correctly classified using the CNN. In FIGURE 6(b), we show examples of wrongly classified ones. We also show the correct labels for these activities. On the second half of the figure (i.e., FIGURE 6(c) and 6(d)), we show examples of sequences of frames and the LSTM predictions for their activities: FIGURE 6(c) shows a correctly classified sequences, and FIGURE 6(d) shows a wrongly classified one alongside with its actual labels.

A preliminary set of experiments was run using a single sensor placed on the ceiling. The results obtained for these experiments were submitted for publication in [21]. Our current work presents more intensive experiments, this time placing the sensor on different positions, and running a more tuned neural network. Nonetheless, this work contains a much deeper comparative study between our work and the conventional ones.

V. DETAILED SYSTEM ARCHITECTURE AND DESCRIPTION

A. DATA COLLECTION

The two sensors kits run the same OS and script to collect the data. However, they collect the data independently from each other. This means that, even though they start simultaneously, a small time difference might occur. In such a case, we synchronize the data later on and discard accordingly a few frames from whichever sensor started before the other. Five people participated in our experiments, each performing different activities for over 5 minutes. Each 5-minute experiment generated over 2000 frames (per sensor), and therefore we collected in total more than 10, 000 frames.

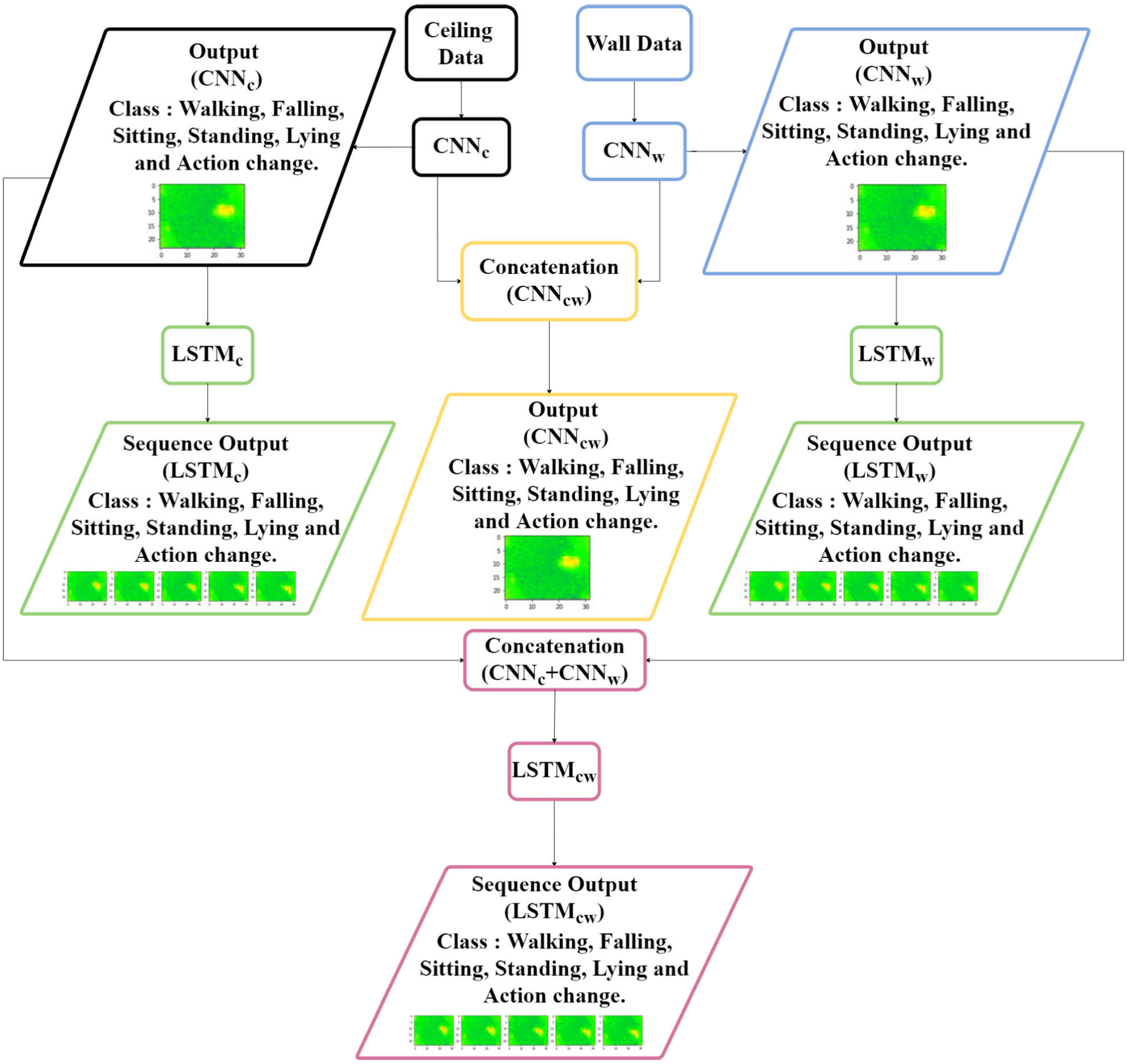

Ceiling Data
Wall Data
Output (CNNc) Class : Walking, Falling, Sitting, Standing, Lying and Action change.
CNNc
CNNw
Output (CNNw) Class : Walking, Falling, Sitting, Standing, Lying and Action change.
Concatenation (CNNcw)
LSTMc
LSTMw
Sequence Output (LSTMc) Class : Walking, Falling, Sitting, Standing, Lying and Action change.
Output (CNNcw) Class : Walking, Falling, Sitting, Standing, Lying and Action change.
Sequence Output (LSTMw) Class : Walking, Falling, Sitting, Standing, Lying and Action change.
Concatenation (CNNc+CNNw)
LSTMcw
Sequence Output (LSTMcw) Class : Walking, Falling, Sitting, Standing, Lying and Action change.

FIGURE 5. A flowchart of the proposed system.

Each 5-minute experiment is referred to as a scenario. The collected 5 scenarios are split into a training data set and a validation data set. The training data set is obviously used to train our deep learning model, whereas the validation data set is used to evaluate the model. Three scenarios are used for training and two scenarios are used for validation.

As stated above, we collected over 10, 000 frames. Frames corresponding to the fractions of time where data are captured by one sensor and not the other, as well as frames where a person is located at the very edge of the coverage area are removed. Table 3 shows the distribution of the remaining frames per activity in both the training and validation sets.

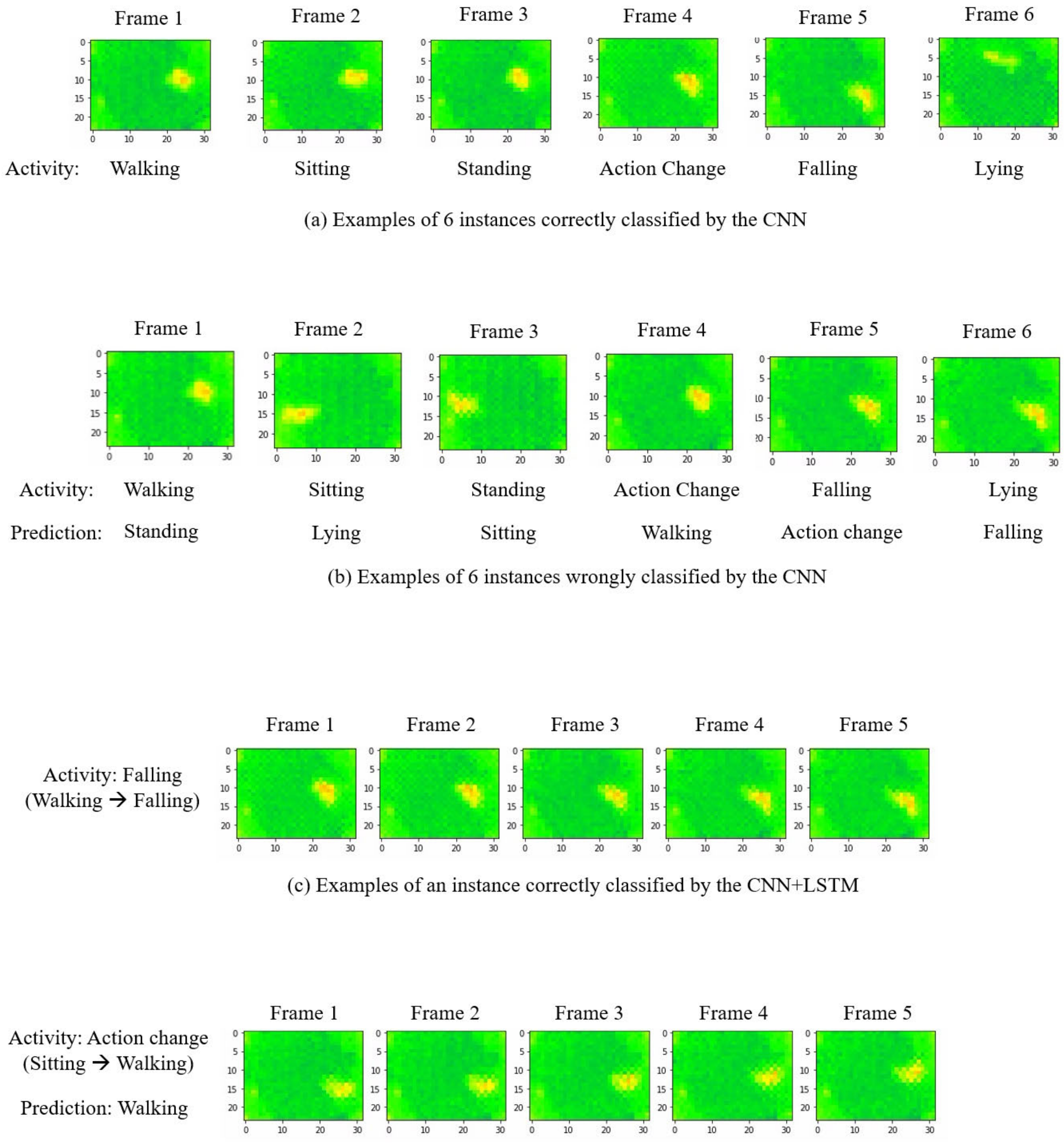
After collecting the data and removing the frames that do not satisfy the requirements of our experiments, an important step needs to be done, which is the annotation of these data. We referred to the images captured by the camera to attribute the activity labels to the sensor frames. FIGURE's. 7 and 8 show some examples of sensor frames alongside with corresponding camera images.

TABLE 3. The frame counts for each activity in the training and the test data sets.

No.	Activity	Train data frames	Test data frames
0	Walking	1282	742
1	Standing	1174	956
2	Sitting	842	726
3	Lying	568	102
4	Action change	371	234
5	Falling	182	156

B. CNN AND LSTM ARCHITECTURE FOR SENSOR DATA CLASSIFICATION

In the first step, we use data collected by each sensor individually to perform the activity detection. Frames collected by the sensor attached to the ceiling are classified using a CNN. Afterwards, the output of the CNN is passed to an LSTM which classifies a sequence of frames for more accurate judgment. In the same way, wall sensor data are classified using CNN and LSTM. The general architecture of the CNN and the



(a) Examples of 6 instances correctly classified by the CNN

(b) Examples of 6 instances wrongly classified by the CNN

(c) Examples of an instance correctly classified by the CNN+LSTM

(d) Examples of an instance wrongly classified by the CNN+LSTM

FIGURE 6. (a) Examples of 6 instances correctly classified by CNN. (b) Examples of 6 instances wrongly classified by CNN. (c) Examples of an instance correctly classified by the CNN+LSTM. (d) Examples of an instance wrongly classified by CNN+LSTM.

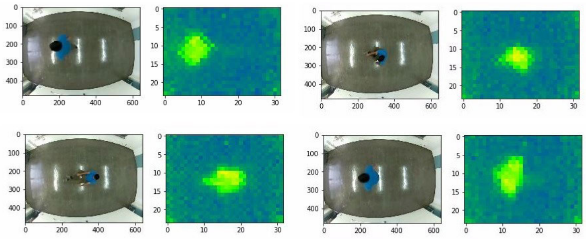


FIGURE 7. Some examples of ceiling sensor frames with their corresponding camera images.

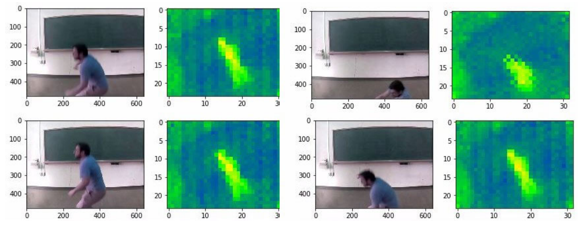


FIGURE 8. Some examples of wall sensor frames with their corresponding camera images.

LSTM is shown in FIGURE 9. Both the ceiling sensor data and the wall sensor data are classified using this architecture.

a: NEURAL NETWORK

As previously stated, throughout this work, we use a hybrid deep learning model to classify the different activities.

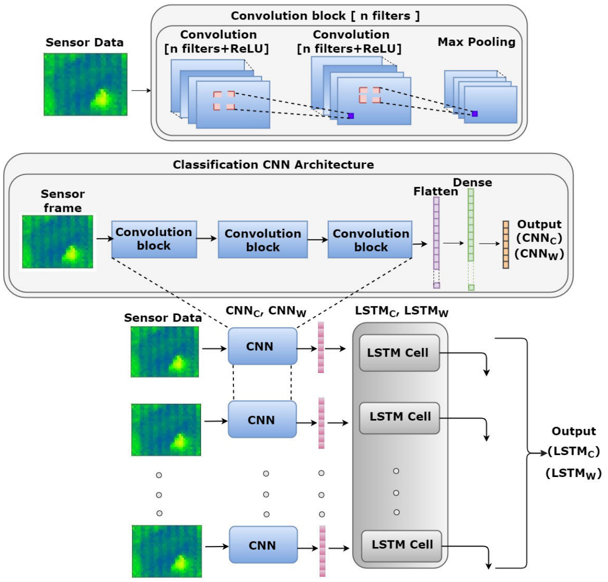


FIGURE 9. The General architecture of the neural network used for classification of both ceiling sensor data and wall sensor data.

Both the CNN and the LSTM networks are composed of the following typical layers:

- *Convolution Layer:* a convolution layer typically takes as input either the raw data or the output of another layer, and applies a set of filters to output more "meaningful" data.
- *Max Pooling Layer:* this layer is usually used to reduce the dimensionality of the features extracted at a given previous layer by picking, for a subset of features, the one with the highest value.
- *Flatten Layer:* this layer is used to flatten the data. In other words, it transforms a multi-dimensional matrix into a single vector.
- *Dense Layer:* it aggregates all the features from the previous layers and maps them to the final features.
- *LSTM Cell:* used for sequential classification of continuous input.

b: CONVOLUTION LAYER

The convolution layers consist of a set of filters with an activation function. The main function of a convolution layer is to get the input data and apply filterers to extract the features. In this CNN architecture, we used 2D-convolution layers. In the current work, we use the term "convolution block" to refer to 2 consecutive 2D-convolution layers with Rectified Linear Unit (ReLU) activation function and filter size 3 × 3, followed by a MaxPooling layer. In our CNN, we have a total of six 2D-convolution layers, where every 2 consecutive layers are followed by a max pooling layer.

c: MAXPOOLING LAYER

The Maxpooling layer function is similar to that of the convolution layer as it also contains filters. It performs a specific function called pooling. MaxPooling is simply taking the maximum value of a subset of values from its input. This operation typicality reduces the dimensionality of the features. In our neural network, we used 2D-max pooling layer with a filter size 2 × 2.

d: FLATTEN LAYER

The flatten layer flattens the previous 2D layers output (which in return is a 2D matrix as well) by converting it into a single vector. This layer has no goal but to connect the 2D output to the fully-connected dense layer that comes after.

e: DENSE LAYER

Dense layers are also referred to in the literature as "fully-connected layers". A dense layer aggregates all the information from the previous layer and maps them into a single feature vector used to identify the activity. The final dense layer outputs the class probability for the different activities. In other words, given an input frame, this last layer outputs a vector whose size is equal to the number of activities, where each value corresponds to the probability of that activity being shown in the frame.

f: LONG SHORT TERM MEMORY(LSTM)

The LSTM is used for sequence classification of input data. It consists of three gates: the input gate, the forget gate and the output gate. LSTM networks can retain information, allowing them to build a more accurate representation of the current state as a function of the previous ones, even ones far away in the past.

g: ACTIVATIONS FUNCTIONS AND HYPERPARAMETERS

In this activity detection system, 2D-Convolution layers use ReLU activation function. This activation function does not activate all the neurons at the same time. Since the output of some neurons is set to zero, only a few neurons are activated making the network sparse, efficient, and easy for computation. The output dense layer uses a softmax function. We use a Stochastic Gradient Decent (SGD) optimizer to optimize the neural network. It reduces the chances of over fitting problem and is less computation-wise costly. For each model, we set Dropout regularization between the layers with a probability equal to 0.2. Batch normalization is used to accelerate the training process. These are the details of the hyper parameters used in all the models of our activity detection system.

Our Neural Networks are designed based on Convolution-LSTM [22] and Siamese Neural Network architecture [23]. This is a common family of neural networks for sequential activity classification. However, the architecture that we propose, as it stands is novel and has been designed taking in mind 3 factors: 1) the type of input data (i.e., sequences of 32 × 24 images) which are very low resolution, 2) the requirement in terms of performance: more complex neural networks might increase the accuracy slightly but not much, and less complex ones have a remarkable performance degradation, and 3) the complexity itself: we expect our model to run on low computation devices such as the

Raspberry Pi (which we used to collect the data). A more complex neural network architecture might end up being very costly for a negligible performance improvement.

C. CNN AND LSTM ARCHITECTURE OF COMBINED SENSOR DATA CLASSIFICATION

The architecture of the CNN used for the classification of the combined data (i.e., data collected from the ceiling sensor and data collected from the wall one) is shown in FIGURE 10. The parameters of the different layers of the neural networks (both the CNN and the LSTM) are the same as explained in the previous subsection. The outputs of the first dense layers of the two sub-networks are concatenated and are connected to a single dense layer whose size is equal to the number of activities. This dense layer obviously outputs the probabilities of the activities.

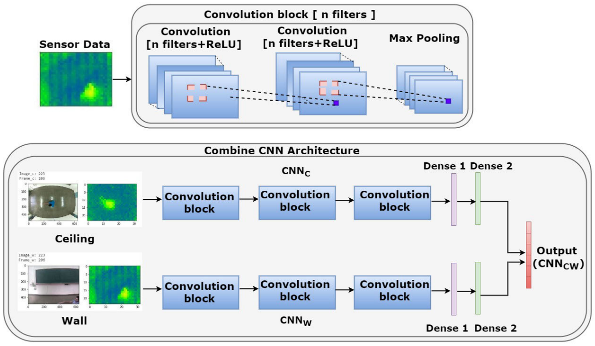


FIGURE 10. The architecture of the combine CNN for classification.

The combined CNN output is passed to the LSTM whose detailed architecture is shown in FIGURE 11. The input to the LSTM is a vector in the time domain whose size is equal to 5. Each time step consists of a vector whose size is equal to 6, which is the output of the CNN.

The combined CNN+LSTM and Combined CNN neural network are designed based on Convolution-LSTM and Siamese neural network, respectively. Our CNN neural network architecture automatically learns the features and the weight of individual frames. The layers of the CNN are then frozen, and the LSTM is trained to use the output of the CNN to run the classification. This has led to a good prediction result when compared to the CNN when used alone. These kinds of architecture have several advantages: On the one hand, the CNN is more robust in classifying imbalanced data, and reach a high accuracy of classification on its own. On the other hand, some activities require observation over an extended period of time to detect the motion. Here the LSTM has a higher potential in detecting such activities.

VI. EXPERIMENTAL RESULTS

We use precision, recall, F1-score, and accuracy as metrics for evaluating the efficiency of the proposed activity detection approach. The True Positive (TP), False Positive (FP), True Negative (TN), and, False Negative (FN) values are reported in the confusion matrix. The evaluation metrics are based on the following formulas:

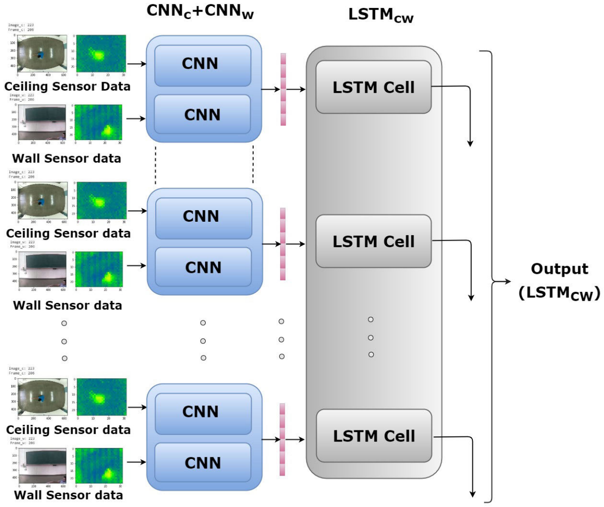


FIGURE 11. The architecture of the combine CNN and LSTM for classification.

$$\text{Precision} = \frac{TP}{TP+FP}, \tag{5}$$

$$\text{Recall} = \frac{TP}{TP+FN}, \tag{6}$$

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{7}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+FP+TN}. \tag{8}$$

We obtained good results from each of the models. However, it is essential to show the capacity for behaviour detection and robustness against false positives. For this, we use precision and recall. Precision measures the correctly classified instances of a given class relative to all the instances classified as belonging to that class. Recall measures the number of correctly classified instances of a given activity relative to all its instances. F1-score is the harmonic mean of both precision and recall.

A. CNN CLASSIFICATION RESULTS

The confusion matrix of the classification of the ceiling sensor data using CNN is shown in Table 4. Based on the observation of this confusion matrix, sitting and standing activities are the most confused ones; and walking and action change activities confusion comes second. From this, we conclude that there is confusion between the sitting and standing activities when classifying ceiling sensor data using CNN.

Next, the performance evaluation for classification of the ceiling sensor data using CNN is shown in Table 5. Similar to our previous observations from the confusion matrix, we can see here that the activities walking and action change are misclassified ones. For instance, despite its high recall, walking activity has low precision. This leads us to believe that the CNN's performance for the walking activity needs

TABLE 4. The confusion matrix of the classification of the ceiling sensor data.

Class	Classified as 0	1	2	3	4	5
Walking-0	**717**	9	0	0	14	2
Standing-1	6	**913**	26	0	11	0
Sitting-2	11	23	**678**	0	14	0
Lying-3	0	7	12	**83**	0	4
Action change-4	19	0	1	6	**208**	0
Falling-3	4	0	3	0	0	**149**

TABLE 5. The precision, recall and F1-score for classification of ceiling sensor data using CNN for each activity.

Activity	Precision	Recall	F1-Measure
Walking	0.95	0.97	0.94
Standing	0.96	0.96	0.96
Sitting	0.94	0.93	0.93
Lying	0.93	0.78	0.85
Action change	0.89	0.94	0.91
Falling	0.96	0.96	0.96

to be improved. Falling and sitting activities have the highest performance, with falling reaching the highest precision and F1.

The results of the classification of the wall sensor data using CNN is shown in Table 6. The confusion matrix shown here, illustrates that this model does not perform well for many activities. Misclassification of the sitting, standing and walking activities is very high owing to the limitations in the detection accuracy arising from the activity being out of the periphery of the wall sensor's range. These limitations need to be overcome for the model to perform well. In addition to the confusion matrix, the detailed performance evaluation of this model is shown in Table 7.

Based on the results so far, it is clear that the detection of some activities such as sitting and falling is better when using the ceiling sensor, whereas the detection of some other activities such as action change is better when using the wall sensor. Clearly, there is a need for improvement of the detection of all the activities in a collective manner. To do so, we combine both the data collected by the ceiling sensor and those collected by the wall sensor and perform the classification on these combined data using CNN.

The results of the classification of the combined data using CNN is presented in Table 8. From these results, we clearly see that the overall misclassification of all the activities has been reduced as compared to the previous results when using individual sensor data for classification using CNN. Walking, standing, sitting, falling, and lying activities have good detection measures compared to the individual sensor data results. The performance evaluation of this classification is shown in Table 9. Walking, standing, and sitting activities show a remarkable improvement in terms of detection based on the precision and recall values.

TABLE 6. The confusion matrix of the classification of the wall sensor data using CNN.

Class	Classified as 0	1	2	3	4	5
Walking-0	**719**	0	0	5	11	7
Standing-1	0	**917**	19	17	3	0
Sitting-2	0	17	**674**	23	12	0
Lying-3	0	12	27	**63**	0	0
Action change-4	12	0	0	1	**217**	4
Falling-5	6	0	2	0	3	**145**

TABLE 7. The precision, recall and F1-score for classification of wall sensor data using CNN for each activity.

Activity	Precision	Recall	F1-Measure
Walking	0.98	0.97	0.97
Standing	0.97	0.96	0.96
Sitting	0.93	0.93	0.93
Lying	0.58	0.62	0.60
Action change	0.93	0.89	0.91
Falling	0.93	0.93	0.93

TABLE 8. The confusion matrix of the classification of the combined sensor(s) data using CNN.

Class	Classified as 0	1	2	3	4	5
Walking-0	**721**	0	0	13	0	8
Standing-1	0	**940**	9	7	0	0
Sitting-2	0	12	**705**	9	0	0
Lying-3	7	4	13	**73**	5	0
Action change-4	5	0	0	7	**220**	2
Falling-5	3	0	0	0	2	**151**

TABLE 9. The precision, recall and F1-score for classification of the combined sensor data using CNN.

Activity	Precision	Recall	F1-Measure
Walking	0.98	0.97	0.97
Standing	0.98	0.98	0.98
Sitting	0.97	0.97	0.97
Lying	0.67	0.72	0.69
Action change	0.94	0.89	0.91
Falling	0.94	0.97	0.95

B. CNN AND LSTM CLASSIFICATION RESULTS

In this section, we discuss the results of the hybrid deep learning model. In this model, the output of the CNN is passed to the LSTM for sequence classification.

The confusion matrix of the sequential classification of the ceiling sensor data is shown in Table 10. From this confusion matrix, it can be inferred that standing and lying are the most misclassified activities. The performance results for this experiment are shown in Table 11. In correlation with the confusion matrix results of this experiment and the prior results relating to the classification of ceiling sensor data using CNN, it can be inferred that lying is the only activity that warrants a further improvement in detection.

The confusion matrix of the classification of the wall sensor data using CNN and LSTM is shown in Table 12.

TABLE 10. The confusion matrix of the classification of the ceiling sensor data using CNN and LSTM.

Class			Classified as			
	0	1	2	3	4	5
Walking-0	**721**	7	0	0	11	3
Standing-1	0	**920**	14	22	0	0
Sitting-2	8	7	**711**	0	0	0
Lying-3	0	5	14	**79**	0	4
Action change-4	6	0	0	0	**221**	7
Falling-3	1	0	0	0	3	**152**

TABLE 11. The precision, recall and F1-score for classification of ceiling sensor data using CNN and LSTM.

Activity	Precision	Recall	F1-Measure
Walking	0.98	0.97	0.97
Standing	0.96	0.96	0.96
Sitting	0.96	0.98	0.97
Lying	0.78	0.77	0.77
Action change	0.94	0.94	0.94
Falling	0.92	0.97	0.94

TABLE 12. The confusion matrix of the classification of the wall sensor data using CNN and LSTM.

Class			Classified as			
	0	1	2	3	4	5
Walking-0	**724**	11	0	0	7	0
Standing-1	0	**923**	24	9	0	0
Sitting-2	7	13	**706**	0	0	0
Lying-3	0	4	26	**72**	0	0
Action change-4	11	0	0	0	**219**	4
Falling-5	5	0	0	0	11	**206**

TABLE 13. The precision, recall and F1-score for classification of wall sensor data using CNN and LSTM.

Activity	Precision	Recall	F1-Measure
Walking	0.97	0.98	0.97
Standing	0.97	0.92	0.94
Sitting	0.93	0.97	0.95
Lying	0.89	0.71	0.79
Action change	0.94	0.89	0.91
Falling	0.97	0.90	0.93

TABLE 14. The confusion matrix of the classification of the combined sensor(s) data using CNN and LSTM.

Class			Classified as			
	0	1	2	3	4	5
Walking-0	**727**	8	0	0	5	2
Standing-1	0	**939**	8	9	0	0
Sitting-2	3	9	**714**	0	0	0
Lying-3	0	0	19	**73**	11	0
Action change-4	13	0	0	0	**219**	2
Falling-5	1	0	0	0	2	**153**

These results in relation to the results obtained from this classification of the wall sensor data using CNN can be used to deduce that lying and sitting activities misclassification rate has been reduced with the improvement in detection for the other activities in case of the current results. Furthermore, Table 13 shows the performance results of the classification using this model. It can be observed the that detection of walking and standing activities performed well in this model. However, detection of lying and sitting activities is still low due to the same reason(s) described in the classification of the wall sensor data using CNN, i.e., the limitation in activity detection due to the subject being out of the sensor's peripheral vision.

Based on the above results obtained so far for sequential classification (CNN+LSTM), certain activities are better detected when using the ceiling sensor data, whereas others are better detected when using the wall sensor data. Therefore, there is still a need for further improvement in the detection performance when using the hybrid model, as our ultimate goal is to have a single model performing well for all the activities. This is best illustrated in the case of detection of lying and sitting activities which is low when the wall sensor data are used compared to when the ceiling sensor is used.

The confusion matrix of the sequential classifier (CNN+LSTM) is presented in Table 14. The results reported in the confusion matrix show an improvement in the detection of all the activities, compared to the case where we used only the CNN. However, lying activity is still relatively poorly detected, requiring further improvement. That being the case, to improve the detection performance of all the activities but particularly that of lying, we combine both the ceiling sensor data and wall sensor data and perform the classification using CNN and LSTM.

The performance evaluation of the classification of the combined sensor data using CNN and LSTM is shown in Table 15. The results indicate that this model performed the best as compared to the other models that have been discussed so far. In particularly noticeable in the case of walking, standing, falling, and action change activities.

TABLE 15. The precision, recall and F1-score for classification of combined sensor(s) data using CNN and LSTM.

Activity	Precision	Recall	F1-Measure
Walking	0.98	0.98	0.98
Standing	0.98	0.98	0.98
Sitting	0.96	0.98	0.97
Lying	0.89	0.71	0.79
Action change	0.94	0.89	0.87
Falling	0.97	0.98	0.97

Accuracy is defined as the correctly classified instances over all the instances of all the activities. We downscale our dataset to 8 × 8 and run it on the previous existing conventional machine learning models. We classified different activities using various models that have been used in the conventional works. Table 16 shows the comparison of classification accuracy for these models. Based on this table, we observe that the combined sensor data classification

TABLE 16. Comparison of the classification accuracy of our models with those in the conventional work.

Methods	No. of the sensor	Position and classification accuracy		
		Ceiling	Wall	Combine ceiling and wall
SVM [13]	1	0.72	-	-
k-NN [14]	1	0.84	-	-
SVM [15]	2	✓	✓	0.90
CNN	2	0.94	0.93	0.96
CNN+LSTM	2	0.96	0.95	0.97

using CNN and LSTM model performed the best and reached over 0.97 accuracy.

VII. DISCUSSION

In the previous section, we presented the details of our experimental results, highlighting the approach we took as we went about designing a hybrid deep learning model to detect all types of activities.

As evident from these results, the classification of the frames using CNN and LSTM individually, with the ceiling sensor and wall sensor data applied separately, had their own respective limitations that were overcome by combining the data from both these sensors.

However, upon comparing these results with existing approaches that tend to use conventional machine learning methodology of manually engineered features, which themselves are speculative on account of being defined by different hypotheses, we proceeded to improve the model further.

To further improve our model, we combined both the CNN and LSTM approaches to build a hybrid approach for sequence activity detection that can quickly identify activities using just 5 frames (<1sec). This is different from the existing methods that use large amounts of sequential frames for detection.

However, the current approach still has a few constraints to be considered and requires solutions to further enhance the activity detection ability and allow for a more general application of the said model. Here are a few of these,

- A certain shortcoming that persisted in the novel hybrid approach relates to specifically detecting lying and action change activities. These could be overcome in the future by using additional sensors that provide for a larger field of view as in contrast to the limited two views available only the single wall and ceiling sensors used currently, thereby enhancing coverage of the subject in a 3-dimensional manner and allowing for better accuracy in the classification of activities.
- Another issue that pertains to this model is the quality of images obtained in the acquisition phase on account of using infrared sensors. The lower resolution of these infrared images can be enhanced by using advanced image processing techniques such as super-resolution and de-noising techniques which would allow for better accuracy in both sequence and individual frame classification.
- The image acquisition can be further affected by the source of infrared radiation and its intensity. This arises from external light sources such as sunlight through the windows, active electronics nearby, pets/other persons in the vicinity, etc. This issue requires a better bounding of the subject with adequate thresholding for a more accurate source when acquiring the images.
- These issues can be further exacerbated when obstacles are present as well. The presence of obstacles by themselves hinders the accurate detection of the subject and necessitates a separate object detection layer to accurately identify the subject. This novel hybrid detection technique is not designed for multiple subjects and our work focuses on elderly people living alone. However, this could also be extended to activity detection of multiple persons in a single field of view for which we need to use localization method [24] to locate the different people, distinguish them one from the other and then perform activity detection.

VIII. CONCLUSION

In this paper, we proposed an activity detection technique using wide-angle low-resolution infrared array sensors. The data collected by the sensors are classified using a hybrid deep learning model. The hybrid deep learning model is designed based on the Convolution-LSTM and Siamese Neural Network architectures. We used two sensors, one placed on the wall and the other placed on the ceiling. This activity detection system involves two phases. In the first phase we classify the wall sensor data and ceiling sensor data using CNN and achieve a classification accuracy of 0.93 and 0.94, respectively. To improve this further, we combined both the sensor data and performed classification using CNN and got an improved accuracy of 0.96. In the second phase, the output of the CNN is passed to an LSTM to achieve better performance. The classification using the ceiling sensor data reaches 0.96 accuracy, whereas that using wall sensor data reaches 0.95 accuracy. When we combine both the wall sensor data and ceiling sensor data, the classification accuracy reaches 0.97. We run some of the existing conventional approaches on our data set and compared the results. Based on these, we can conclude that by combining the data collected by the sensor placed on the ceiling and that placed on the wall, and using CNN and LSTM, we get the highest classification accuracy which is 0.97.

REFERENCES

[1] *Department of Economic and Social Affairs, Population Division (2019)*, United Nations, World Population Ageing, Highlights (ST/ESA/SER.A/430), New York, NY, USA, 2019.

[2] Government of Japan. *Annual Report on the Aging Society*. [Online] Available: https://www8.cao.go.jp/kourei/english/annualreport/2019/pdf/2019.pdf

[3] T. Le Nguyen and T. T. H. Do, "Artificial intelligence in healthcare: A new technology benefit for both patients and doctors," in *Proc. Portland Int. Conf. Manage. Eng. Technol. (PICMET)*, Aug. 2019, pp. 1–15.

[4] Y. Song and T. J. M. van der Cammen, "Electronic assistive technology for community-dwelling solo-living older adults: A systematic review," *Maturitas*, vol. 125, pp. 50–56, Jul. 2019.

[5] E. Bulbul, A. Cetin, and I. A. Dogru, "Human activity recognition using smartphones," in *Proc. 2nd Int. Symp. Multidisciplinary Stud. Innov. Tech. (ISMSIT)*, Ankara, pp. 1–6, Oct. 2018.

[6] F. Shahmohammadi, A. Hosseini, C. E. King, and M. Sarrafzadeh, "Smartwatch based activity recognition using active learning," in *Proc. IEEE/ACM Int. Conf. Connected Health, Appl., Syst. Eng. Technol. (CHASE)*, Jul. 2017, pp. 321–329.

[7] S. Balli, E. A. Sağbaş, and M. Peker, "Human activity recognition from smart watch sensor data using a hybrid of principal component analysis and random forest algorithm," *Meas. Control*, vol. 52, nos. 1–2, pp. 37–45, Jan. 2019.

[8] A. Eerdekens, M. Deruyck, J. Fontaine, L. Martens, E. D. Poorter, and W. Joseph, "Automatic equine activity detection by convolutional neural networks using accelerometer data," *Comput. Electron. Agricult.*, vol. 168, Jan. 2020, Art. no. 105139.

[9] A. S. A. Sukor, A. Zakaria, and N. A. Rahim, "Activity recognition using accelerometer sensor and machine learning classifiers," in *Proc. IEEE 14th Int. Colloq. Signal Process. Appl. (CSPA)*, Mar. 2018, pp. 233–238.

[10] M. L. Gavrilova, Y. Wang, F. Ahmed, and P. Polash Paul, "Kinect sensor gesture and activity recognition: New applications for consumer cognitive systems," *IEEE Consum. Electron. Mag.*, vol. 7, no. 1, pp. 88–94, Jan. 2018.

[11] Y. Hino, J. Hong, and T. Ohtsuki, "Activity recognition using array antenna," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 507–511.

[12] J. Hong, S. Tomii, and T. Ohtsuki, "Cooperative fall detection using Doppler radar and array sensor," in *Proc. IEEE 24th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Sep. 2013, pp. 3492–3496.

[13] S. Mashiyama, J. Hong, and T. Ohtsuki, "A fall detection system using low resolution infrared array sensor," in *Proc. IEEE 25th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Sep. 2014, pp. 2109–2113.

[14] S. Mashiyama, J. Hong, and T. Ohtsuki, "Activity recognition using low resolution infrared array sensor," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 495–500.

[15] K. Kobayashi, T. Ohtsuki, and K. Toyoda, "Human activity recognition by infrared sensor arrays considering positional relation between user and sensors," in *Proc. Smart City Based Ambient Intelliegnce*, 2018, pp. 1–6. [Online]. Available: https://ipsj.ixsq.nii.ac.jp/ej/index.php?active_action=repository_view_main_item_detail&page_id=13&block_id=8&item_id=178209&item_no=1

[16] X. Fan, H. Zhang, C. Leung, and Z. Shen, "Robust unobtrusive fall detection using infrared array sensors," in *Proc. IEEE Int. Conf. Multisensor Fusion Integr. Intell. Syst. (MFI)*, Nov. 2017, pp. 194–199.

[17] Y. Taniguchi, H. Nakajima, N. Tsuchiya, J. Tanaka, F. Aita, and Y. Hata, "A falling detection system with plural thermal array sensors," in *Proc. Joint 7th Int. Conf. Soft Comput. Intell. Syst. (SCIS) 15th Int. Symp. Adv. Intell. Syst. (ISIS)*, Dec. 2014, pp. 673–678.

[18] C. Taramasco, T. Rodenas, F. Martinez, P. Fuentes, R. Munoz, R. Olivares, V. H. C. De Albuquerque, and J. Demongeot, "A novel monitoring system for fall detection in older people," *IEEE Access*, vol. 6, pp. 43563–43574, 2018.

[19] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine," in *Proc. 4th Int. Workshop Ambient Assist. Living Home Care*, Vitoria-Gasteiz, Spain, Dec. 2012, pp. 216–223.

[20] A. Mannini and A. M. Sabatini, "Machine learning methods for classifying human physical activity from on-body accelerometers," *Sensors*, vol. 10, no. 2, pp. 1154–1175, Feb. 2010.

[21] K. A. Muthukumar, M. Bouazizi, and T. Ohtsuki, "Activity detection with a wide-angle low-resolution infrared array sensor using deep learning," in *Proc. IEEE EMBC*, to be published. [Online]. Available: https://embc.embs.org/2021/

[22] J. Donahue, L. A. Hendricks, M. Rohrbach, S. Venugopalan, S. Guadarrama, K. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 677–691, Apr. 2017, doi: 10.1109/TPAMI.2016.2599174.

[23] S. Berlemont, G. Lefebvre, S. Duffner, and C. Garcia, "Class-balanced siamese neural networks," *Neurocomputing*, vol. 273, pp. 47–56, Jan. 2018.

[24] M. Bouazizi and T. Ohtsuki, "An infrared array sensor-based method for localizing and counting people for health care and monitoring," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 4151–4155.

MUTHUKUMAR K. A. (Graduate Student Member, IEEE) received the Bachelor of Technology degree in computer science engineering from the National Institute of Technology Tiruchirappalli (NIT-Trichy), India, in 2012, and the Master of Technology degree in computer science engineering from Pondicherry University, India, in 2016. He is currently pursuing the Ph.D. degree with Keio University, Japan. He enrolled as a Research Student with Keio University, in 2018.

MONDHER BOUAZIZI (Member, IEEE) received the bachelor's degree in communications from SUPCOM, Carthage University, Tunisia, in 2010, and the master's and Ph.D. degrees from Keio University, in 2017 and 2019, respectively. He worked as a Telecommunication Engineer (access network quality and optimization) for period of three years with Ooredoo Tunisia (Ex. Tunisiana). He is currently working as a Special Assistant Professor with Keio University. He is also a Student Member of IEICE.

TOMOAKI OHTSUKI (Senior Member, IEEE) received the B.E., M.E., and Ph.D. degrees in electrical engineering from Keio University, Yokohama, Japan, in 1990, 1992, and 1994, respectively.

From 1993 to 1995, he was a Special Researcher of Fellowships of the Japan Society for the Promotion of Science for Japanese Junior Scientists. From 1994 to 1995, he was a Postdoctoral Fellow and a Visiting Researcher in electrical engineering with Keio University. From 1995 to 2005, he was with the Science University of Tokyo. From 1998 to 1999, he was with the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA, USA. He joined Keio University, in 2005. He is currently a Professor with Keio University. He is engaged in research on wireless communications, optical communication, signal processing, and information theory. He has published more than 200 journal articles and 410 international conference papers. He is a Fellow of the IEICE. He was a recipient of the 1997 Inoue Research Award for Young Scientist, the 1997 Hiroshi Ando Memorial Young Engineering Award, the Ericsson Young Scientist Award 2000, the 2002 Funai Information and Science Award for Young Scientist, the IEEE the 1st Asia-Pacific Young Researcher Award 2001, the 5th International Communication Foundation (ICF) Research Award, the 2011 IEEE SPCE Outstanding Service Award, the 27th TELECOM System Technology Award, ETRI Journal's 2012 Best Reviewer Award, and the 9th International Conference on Communications and Networking in China 2014 (CHINACOM'14) Best Paper Award. He served as the Chair for IEEE Communications Society, Signal Processing for Communications and Electronics Technical Committee. He has served as the General-Co Chair, Symposium Co-Chair, and TPC Co-Chair of many conferences, including IEEE GLOBECOM 2008, SPC, IEEE ICC2011, CTS, IEEE GCOM2012, SPC, IEEE ICC2020, SPC, IEEE APWCS, IEEE SPAWC, and IEEE VTC. He was a Vice President and the President of Communications Society of the IEICE. He served as a Technical Editor for the *IEEE Wireless Communications Magazine* and an Editor for *Physical Communications* (Elsevier). He is also serving as an Area Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and an Editor for the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS. He gave tutorials and keynote speech at many international conferences, including IEEE VTC and IEEE PIMRC. He is a Distinguished Lecturer of the IEEE Vehicular Technology Society.