Robust Unobtrusive Fall Detection using Infrared Array Sensors

Xiuyi Fan\textsuperscript{1}, Huiguo Zhang\textsuperscript{1}, Cyril Leung\textsuperscript{2}, Zhiqi Shen\textsuperscript{3}

\textbf{Abstract}—As the world’s aging population grows, fall is becoming a major problem in public health. It is one of the most vital risk to the elderly. Many technology based fall detection systems have been developed in recent years with hardware ranging from wearable devices to ambience sensors and video cameras. Several machine learning based fall detection classifiers have been developed to process sensor data with various degrees of success. In this paper, we present a fall detection system using infrared array sensors with several deep learning methods, including long-short-term-memory and gated recurrent unit models. Evaluated with fall data collected in two different sets of configurations, we show that our approach gives significant improvement over existing works using the same infrared array sensor.

I. INTRODUCTION

Aging is a global challenge faced by many countries in the world. The rapid growth of the aging population puts high demand for relevant assistive technologies supported by various sensor-actuator systems [1]. There are various types of sensors utilized in assisted living, including cameras [2], light sensors, accelerometers [3], temperature sensors, gyroscope, barometer, infrared sensors [4], etc. These sensors are rich data sources for understanding various aspects of a user’s daily life, ranging from health and fitness monitoring, personal biometric signature, navigation and localization [5]. In this context, one particular problem is the detection of fall of the elderly. Fall is the most vital risk to the elderly’s health as over one in every three elderly people suffer from fall consequences [6], [7]. In event of fall, it is urgent to provide immediate treatment of the injured. Thus the quick detection of fall is essential for on time treatment [8].

Technology based fall detection has been of great interest. It has generated a wide range of applied research and has prompted the development of telemonitoring systems to enable the early diagnosis of fall conditions [9]. Mubashir et. al. distinguish fall detection systems into three categories, wearable devices, ambience sensors and cameras [5]. The first category needs the subject of interest wearing a wearable device all the time whereas the last two only to deploy the device in the vicinity of the subject.

In addition to sensor development, different data classification techniques have been developed for fall detection. From raw sensor data, various data processing algorithms have been proposed in the literature. Roughly speaking, there are two schools of methods for fall detection: rule-based methods that detect falls with domain knowledge and machine learning based approaches “learn fall characteristics” from training data [1], [9].

In this work, we present a fall detection system that is based on data collected from Grid-Eye Infrared Array Sensors, which are low cost, low resolution infrared thermal image temperature sensors. These low resolution sensors have less intrusion of privacies when compared with high resolution sensors like RGB cameras. Sensor data is processed with several mainstream deep learning models, including the long short term memory (LSTM) [10] and gated recurrent unit (GRU) models [11]. We have also experimented these models with attention mechanisms as proposed in [12]. We compare our approaches with the fall detection system reported in [13], which also uses the same Grid-Eye sensor, and we show that our approach yields improvement over existing ones.

The rest of this paper is organized as follows. Section II introduces several existing works on fall detection. Section III introduces deep learning classifiers we developed in this work. Section IV presents performance evaluation of the developed fall detection system. We conclude the paper and discuss future research directions in section V.

II. RELATED WORK

Existing fall detection systems can be categorized into three types, wearable devices, camera systems and ambience sensors [5]. Wearable devices are sensors attached to a human body to collect body movements and to recognize activities. Most wearable devices use accelerometers and gyroscopes [14], [15]. In these fall detection systems, sensors are attached to different parts of the user’s body such as waist [7], chest [6], and shoes [16]. One major problem with wearable device based methods is that the user has to wear the device all the time, which causes a great amount of inconvenience. Also, users often forgot to wear such devices from time to time.

Camera based fall detection systems normally use RGB cameras [17]. Recently, several studies also use Microsoft Kinect [18], [19]. Camera-based devices are commonly deployed through the elderly’s house or at public places. There are two limitations with these systems, privacy intrusion with video monitoring and the lack of system robustness.

Ambience sensor based fall detection systems have also been studied. Different sensors or devices such as doppler radar [20], passive infrared sensors [21], [22], [13], [23],
pressure sensors [24], [25], sound sensors [26] and Wi-Fi routers [27] have been tested for fall detection.

Many research has been devoted to the study of fall detection classification algorithms [8], [28]. There are mainly two categories of methods developed, rule-based methods that depend much on domain knowledge and machine learning methods that recognize fall characteristics from sensor data [1], [9]. For instance, [29], [30], [31], [32] are some early fall detection works with threshold-based algorithms. In those works, thresholds are set such that if any of these thresholds is exceeded, then a fall alert is triggered. The major drawback of these approaches is the lack of adaptability and flexibility.

At the same time, various machine learning based fall detection classifiers have been developed [33]. Mainstream machine learning approaches, including decision trees [34], support vector machines (SVM) [35], k-nearest neighbours (k-NN) [36] and hidden Markov models [37] have been applied in fall detection, see e.g., [38], [39], [40], [23]. Many of these approaches rely on manually designed features for classification.

The following works are most relevant to ours. L. Liu et. al. [20] develop a dual Doppler radar system for fall detection. A fusion methodology combines partial decision information from two sensors in three different classifiers, k-NN, SVM and Bayes to form a fall/non-fall decision based on Mel-frequency Cepstral Coefficients (MFCC) features. Its performance measured with AUC is 0.88 and 0.97.

Liu et. al. [21] propose a two-layer hidden Markov model for recognizing a fall event based on the signals of five passive infrared sensors which were placed at different heights on the wall. The associated sensitivity and specificity of the falls algorithm were 92.5% and 93.7%, respectively.

Chen et. al. [23] use 16-by-4 thermopile array sensors for fall detection and elderly tracking. Two sensors are used in their system with a k-NN classifier. They have reached 95.25% sensitivity, 90.75% specificity and 93% accuracy in their experiment. Sixsmith and Johnson [41] developed a Smart Inactivity Monitor using array-based detectors which also detects falls.

Mashiyama et. al. [13] propose a system of fall detection using an infrared array sensor. From a data sequence obtained in a fixed window, four manually crafted features, number of consecutive frames, maximum number of pixels, maximum variance of temperature and distance of a maximum temperature pixel, are extracted from the sequence and used to classify falls or non-falls using the k-NN algorithm. Experiment results with their testing data show that their system reaches 94% accuracy.

III. FALL DETECTION CLASSIFIERS

At the core of our fall detection system is the infrared array sensor, Grid-Eye (AMG8832). A Grid-Eye sensor outputs an 8-pixel by 8-pixel temperature distribution in its 60-degree field of view at a maximum 10-frame per second rate. Its maximum detection distance is 5m if there is a $\geq 4°C$ temperature difference between the foreground object and the background ambience. We use a ZigBee CC2530 as a microprocessor to control the sensor via an I$^2$C bus as shown in Figure 1. The measured temperature distribution is sent to another ZigBee CC2530 at a 10Hz rate. A standard PC is then used for data processing and classification.

![Fig. 1. The Grid-Eye sensor package used in our experiment.](image-url)

Although a Grid-Eye sensor measures temperature in a large range (-20°C to 100°C), its temperature accuracy is only 3.0°C. Since thermal image based fall detection depends on correctly identifying the abrupt movement of a human body, the ability to recognize the subtle temperature difference between the human body and the ambience is the key to ensure correct detections. However, as illustrated in Figure 2, data obtained from Grid-Eye sensors is noisy. (In this figure, warm colour indicates high temperature.) Thus, we develop a fall detection system with two main components: (1) data filters for pre-processing and (2) neural networks for classification. As illustrated in Figure 3, data produced by the Grid-Eye is firstly filtered with one of the filters. Filtered data is then passed to neural network classifiers.

Three filters, Median, Gaussian and Wavelet, have been experimented in this work. For neural network classifiers, we have experimented with two-layer perceptron networks (Figure 4), long short-term memory (LSTM) networks and gated recurrent unit (GRU) networks (Figure 5), each with and without attention links.

As illustrated in Figure 6, the developed system works as follows. At each time step $t$, the Grid-Eye outputs thermal reading represented with a $1 \times 64$ vector. To detect fall, we examine data collected in a 2-second (outer) window. Since the Grid-Eye is running at 10Hz, 20 $1 \times 64$ vectors are collected during each (outer) window. We then filter data...
stored in this outer window with one of the three filters. For both median and Gaussian filters, an inner window of size 5 is used. For the wavelet filter, we use Daubechies 4 tap wavelet. The filtering process does not change the size of the data. Filtered data is then sent to neural networks for classification.

Two-layer perceptron networks with the following configuration are selected for their simplicity. The input layer contains \(64 \times 20 = 1280\) nodes (64 is the length of the Grid-Eye output vector and 20 is the size of the outer window). The fully connected hidden layer contains 400 nodes. The output layer contains 2 nodes (indicating a fall and not a fall, respectively).

LSTM and GRU networks have seen many success in recent years. They both contain “memory structures”, i.e., LSTM cells and GRU units, to store past information. As illustrated in Figure 5, the input layers of our LSTM and GRU networks both contain 64 nodes. There is a fully connected perceptron layer with 64 nodes between the LSTM / GRU layer and the 2-node output layer. The LSTM model can be described with the following equations:

\[
\begin{align*}
    i & = \sigma(x_t U^i + s_{t-1} W^i) \\
    f & = \sigma(x_t U^f + s_{t-1} W^f) \\
    o & = \sigma(x_t U^o + s_{t-1} W^o) \\
    g & = \tanh(x_t U^g + s_{t-1} W^g) \\
    c_t & = c_{t-1} \circ f + g \circ i \\
    s_t & = \tanh(c_t) \circ o
\end{align*}
\]

Here, \(\sigma\) is the sigmoid function. \(\circ\) denotes element-wise multiplication. \(x_t\) is the input at time \(t\), \(s_t\) is the output of the cell at time \(t\). \(U_s\)s and \(W_s\)s are weight matrices connecting various components. Specifically, in our system, \(x_t\) is a 1-by-64 vector; \(s_t\) is a 1-by-64 vector; \(U_s\)s, are 64-by-64 matrices; \(W_s\)s are 64-by-64 matrices.

GRU [11] is a recently proposed variation of the LSTM model. The main difference is that, instead of using three gates to control memory updates, a GRU unit uses only two gates. Formally, a GRU model can be described with the following equations:

\[
\begin{align*}
    z & = \sigma(x_t U^z + s_{t-1} W^z) \\
    r & = \sigma(x_t U^r + s_{t-1} W^r) \\
    h & = \tanh(x_t U^h + (s_{t-1} \circ r) W^h) \\
    s_t & = (1 - z) \circ h + z \circ s_{t-1}
\end{align*}
\]

Again, \(\sigma\) is the sigmoid function. \(x_t\) is the input at time \(t\), \(h\) is the output. \(s_t\) is the internal state of a GRU unit at time \(t\). The size of \(U_s\)s and \(W_s\)s are the same as in LSTM. Essentially, we use the same network structure as our LSTM implementation, with LSTM cells replaced by GRU units.

Introducing attention mechanism into both LSTM and GRU models in this work is very simple. Conceptually, the attention mechanism provides a means for specifying the relative importance of each frame in a classification window (20-frames in our case). For instance, \(s_i\) in Equation 6 for \(t = 20\) not only depends on \(s_{i9}\) but also (directly) depends on all previous \(s_j\), for \(1 \leq i \leq 19\), i.e.,

\[
s_{20} = \sum_{0 \leq i < 20} \omega_i s_i,
\]

for some \(\omega_i\) also learned with backward propagation though time as \(U\) and \(W\).

IV. PERFORMANCE EVALUATION

To evaluate the performance of the developed system, we conduct fall detection experiment in our laboratory environment (Figure 7). In our test, we have created a dataset with 312 falls in two sets of configurations. As illustrated in Figure 8, in the first set of experiments, the testing subject falls perpendicular to the Grid-Eye sensor at A, B and C three different positions. In the second set of experiments, the testing subject falls parallel to the Grid-Eye sensor, also at A, B and C three different positions. In both configurations, negative examples including randomly walking in the room, slowly sitting down, jumping, running and laying down in front of the sensor have been performed. The dataset has been created in multiple sessions crossing several days with ambient temperature ranging from 19°C to 23°C.

For evaluation, we have divided the dataset into a training set with 240 falls and a testing set with 72 falls with each falling position contains exactly the same number of falls. Since robust fall detection requires high ratings in both precision and recall, reducing both false positives and false negatives, we compare results with F1 scores for each test.
case, defined as follows.

\[
\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}},
\]

\[
\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}},
\]

\[
F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.
\]

Experiment results from our systems are shown in Figure 9 (for falls perpendicular to the sensor) and Figure 10 (for falls parallel to the sensor). Overall, we make the following observations.

- Measured by F1 scores, all classifiers perform better in settings where users fall parallelly to the sensor. This indicates that falling-parallel-to-the-sensor is intrinsically easier to classify than falling-perpendicular-to-the-sensor.
- Introducing filters specifically to remove noise improves the performance in general. Amongst three filters tested, the simple median filter performs better than the other two.
- There is no clear winner between LSTM models and GRU models. The memory ability of both models works well.
- Introducing attention mechanisms in both LSTM and GRU models does not consistently improve the performance. This may suggest that fall detection takes...
information from all frames containing a fall equally and it gives no advantage to focus the detection at any single moment of the fall.

- When the classification problem is easy (parallel settings), MLP does not expose its weakness; however, when the problem gets more difficult (perpendicular settings), models explicitly recording previous information perform significantly better.

Fig. 9. F-scores. Testing subject fall perpendicular to the sensor.

In order to put our results into perspective, we compare our approaches with the model presented in [13], which uses the same Grid-Eye sensor with a k-NN classifier with four manually crafted features. We replicate their system and tested on our dataset, the comparison results are shown in Table I (perpendicular to the sensor) and II (parallel to the sensor). From these two tables, we see that their approach also performs better when falls are parallel to the sensor. However, overall, their k-NN classifier with manually crafted features performs worse than any of our neural network based approaches with data filtering.

V. CONCLUSION

Fall is a major health threat to the elderly. In event of fall, it is urgent to provide immediate treatment to the injured people. In this paper, we present a fall detection system using Grid-Eye infrared array sensor. Due to its low spatial resolution, infrared array sensor incurs little privacy intrusion and can be deployed to sensitive areas such as washrooms, which are known to be fall-prone. For data processing, we have taken a two-step approach: (1) pre-processing data filtering and (2) machine learning classification with neural networks. For filtering, we have experimented with Wavelet, Gaussian and Median filters. For classification, we have experimented with several deep learning models, including multi-layer perceptrons, LSTM and GRU. To evaluate our approaches, we have created a dataset containing over 300 falls in multiple configurations. We then compare our work with an existing work using the same infrared array sensor but with different classification techniques and show significantly improved classification accuracy. In the future, we would like to (1) perform in depth theoretical study, including computational complexity analysis, of the proposed methods, (2) deploy our system to nursing homes for real-world experiment and (3) explore fall detection with other ambience sensor systems and deployment configurations.

### Table I: Fall Detection Performance (Falls are perpendicular to the Grid-Eye).

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRU-ATT</td>
<td>0.75</td>
<td>0.83</td>
<td>0.79</td>
</tr>
<tr>
<td>GRU</td>
<td>0.82</td>
<td>0.72</td>
<td>0.94</td>
</tr>
<tr>
<td>LSTM-ATT</td>
<td>0.72</td>
<td>0.94</td>
<td>0.82</td>
</tr>
<tr>
<td>LSTM</td>
<td>0.85</td>
<td>0.82</td>
<td>0.89</td>
</tr>
<tr>
<td>MLP</td>
<td>0.67</td>
<td>0.72</td>
<td>0.69</td>
</tr>
<tr>
<td>k-NN [13]</td>
<td>0.52</td>
<td>1</td>
<td>0.68</td>
</tr>
</tbody>
</table>

### Table II: Fall Detection Performance (Falls are parallel to the Grid-Eye).

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRU-ATT</td>
<td>0.97</td>
<td>1</td>
<td>0.99</td>
</tr>
<tr>
<td>GRU</td>
<td>0.95</td>
<td>1</td>
<td>0.97</td>
</tr>
<tr>
<td>LSTM-ATT</td>
<td>0.95</td>
<td>1</td>
<td>0.97</td>
</tr>
<tr>
<td>LSTM</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>MLP</td>
<td>0.97</td>
<td>1</td>
<td>0.99</td>
</tr>
<tr>
<td>k-NN [13]</td>
<td>0.83</td>
<td>0.97</td>
<td>0.9</td>
</tr>
</tbody>
</table>

### References


