Wi-Fi-CSI-based Fall Detection by Spectrogram Analysis with CNN

Takashi Nakamura*, Mondher Bouazizi†, Kohei Yamamoto*, and Tomoaki Ohtsuki†
* Graduate School of Science and Technology Keio University
3-14-1 Hiyoshi, Kohoku, Yokohama, Kanagawa 223-8522, Japan
Email: t.nakamura@ohtsuki.ics.keio.ac.jp, yamamoto@ohtsuki.ics.keio.ac.jp
† Department of Information and Computer Science Keio University
3-14-1 Hiyoshi, Kohoku, Yokohama, Kanagawa 223-8522, Japan
Email: bouazizi@ohtsuki.ics.keio.ac.jp, ohtsuki@ics.keio.ac.jp

GLOBECOM 2020 - 2020 IEEE Global Communications Conference | 978-1-7281-8298-8/20/$31.00 ©2020 IEEE | DOI: 10.1109/GLOBECOM42002.2020.9322323

***Abstract*—Fall detection system has a great demand for elderly people living alone. Wi-Fi CSI (Channel State Information) based fall detection method can be used to build non-intrusive and non-space-limited fall detection systems. In the conventional work on Wi-Fi CSI based fall detection, a classification performance degradation has been observed when data in different environments is used for learning and testing data. Also, that method can not capture accurate features of motion due to the signal distortion during the noise reduction, and it can not segment signals accurately when the SNR (Signal to Noise power Ratio) is small. In this paper, we propose a spectrogram image-based fall detection using Wi-Fi CSI. Unlike the conventional method, CSI is segmented with a certain sliding time window, and then the classifier detects fall by using the spectrogram image generated from segmented CSI. We use a CNN (Convolutional Neural Network) for binary classification of the spectrogram images of the fall and non-fall motions. We carried out experiments to evaluate the classification performance of our proposed method against the conventional one by using motion data in two different rooms for learning and testing data. As a result, we confirmed that our proposed method outperformed conventional one and reached 0.90 accuracy.**

I. Introduction

Falling is a critical accident for elderly people living alone. It can lead to serious injury, disability, or death. World Health Organization reports that 28 – 35% people who are over 65 years old fall at least once a year [1]. When such event happens, it is important to rescue the elder person in question immediately [2]. Thus, it is highly demanded to detect fall, in particular for those who are living alone.

To detect a fall, various methods have been proposed with a wearable sensor [3], [4], or a camera [5], [6]. However, the wearable sensor-based methods need to attach sensors on a subject's body, which is not preferred in practical systems. Also, the camera-based methods suffer from the privacy invasion, since this method uses the image information. Due to the aforementioned limitations, the uses of an FMCW (Frequency Modulated Continuous Wave) radar [7], an infrared sensor [8], and Wi-Fi have been also investigated. The FMCW radar-based method detects a fall by extracting features from the radar reflection signals using the time-frequency analysis [7]. The infrared sensor-based method detects a fall based on the temperature change between adjacent frames [8]. Both the FMCW radar and infrared sensor-based methods need to fix devices somewhere in the room, e.g., wall, ceiling, and furniture. However, the coverage area is not large enough to detect fall in a room with one FMCW radar or one infrared sensor. In contrast, the Wi-Fi-based fall detection uses off-the-shelf Wi-Fi modules. Furthermore, thanks to the multipath effects of Wi-Fi signals, the coverage area of the Wi-Fi-based fall detection method is wide, compared with the other existing ones [9]. Therefore, the Wi-Fi-based method can be used to build an non-intrusive and non-space-limited fall detection system.

In general, the Wi-Fi-based fall detection method uses CSI (Channel State Information). CSI delivers signal's amplitude and phase informations over multiple channels and OFDM subcarriers. By analyzing the change of CSI, it is possible to detect human motions. WiFall [10] has been proposed as one of the Wi-Fi CSI-based fall detection methods. This method extracts time domain features from the amplitude information of CSI. Also, RT-Fall [11], which is based on WiFall, extracts time domain features from not only the amplitude information of CSI but also the phase information of the CSI. However, WiFall and RT-Fall are not robust to the environment change. If the model, which is trained with a certain room, is applied in another room, the classification performance degrades, which indicates that the extracted features depend on the environment. To deal with this problem, FallDeFi has been proposed [12], which is CSI-based fall detection using time-frequency domain features of CSI. FallDeFi calculates a spectrogram from CSI and extracts features from the spectrogram such as event duration, spectral entropy, and so on. Although FallDeFi tackles the environmental robustness, it still has the problem of signal segmentation due to temporal signal noise or environment change. Specifically, FallDeFi uses thresholds to detect fall candidate signals, and these are optimized for each room to remove noise energy. If the sum of spectral power in a certain frequency range is above the threshold, the signal is segmented as the fall candidate signal. However, if the SNR (Signal to Noise power Ratio) of the signal is small, the classification performance degrades. This is because noise reduction during spectrogram processing removes weak energy part of the motion, which leads to the extraction of

978-1-7281-8298-8/20/$31.00 ©2020 IEEE

non-accurate features. In addition, there is a possibility that the signal of the motion candidate is extracted even if there is no motion.

In this paper, we propose a spectrogram image-based fall detection using Wi-Fi CSI. Unlike FallDeFi, in the proposed method, CSI is segmented with the sliding time window, and then the fall is detected based on the segmented CSI for each time window. More specifically, a spectrogram is calculated from the segmented CSI, and the fall is detected by inputting the segmented CSI into a CNN (Convolutional Neural Network) with a binary classification output. CNN is one of the deep learning architectures that have been proven to be very successful in various fields, e.g., the image recognition [13] and the sound recognition [14]. By using CNN, our method classifies the spectrogram based on the features that are not dependent on the environment.

We conducted experiments in two different rooms to classify a fall and some non-fall motions. We collected data by using commodity Wi-Fi devices. We then evaluated the classification performance, and compared our results to those of FallDeFi. As a result, we confirmed that our proposed method outperformed FallDeFi and reached 0.90 accuracy.

The organization of this paper is as follows. We first explain the fundamental of Wi-Fi CSI in Section II. In Section III, we summarize conventional work on fall detection using Wi-Fi CSI. In Section IV, our proposed method is described. The evaluation setup and the result are shown in Section V. Finally, we conclude this paper in Section VI.

II. Wi-Fi Channel State Information

In this section, for a better understanding of the proposed method, we explain the Wi-Fi CSI. When a transmitter (Tx) and a receiver (Rx) are inside of an ordinary room, the transmitted signal propagates through different paths, which is modeled as a wireless channel. A wireless channel is modeled as Eq. (1),

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{N}, \tag{1}$$

where $\mathbf{Y}$ and $\mathbf{X}$ are the received and transmitted signal vector, respectively, $\mathbf{N}$ is a noise vector, and $\mathbf{H}$ is the matrix of channel frequency response values which by complex numbers. CSI reflects $\mathbf{H}$ for each time. CSI value is captured for each channel (i.e., antenna pair) and each OFDM subcarrier. The number of CSIs N_h is expressed as Eq. (2),

$$N_h = N_t * N_r * N_c, \tag{2}$$

where N_t and N_r are the number of transmit and receive antennas, respectively, and N_c is that of subcarriers.

When a transmitter and a receiver are inside of an ordinary room, the transmitted signals are reflected by some objects, such as a ceiling, a floor, a furniture, or a human body. If a subject moves, the CSI changes with some patterns specific to the motion. By analyzing the change of CSI value, it possible to detect human motions.

III. Related Work

In this section, we describe the conventional work related with the Wi-Fi-based fall detection.

A. Time-Domain-Based Method

WiFall [10] is the first work on fall detection using Wi-Fi CSI. WiFall segments signals into a certain time length by capturing CSI changes with Local Outlier Factors (LOF), and extracts time-domain features from the amplitude information of the segmented CSI. In the experiment, the pre-defined four motions are measured and are classified into fall and others by using one-class Support Vector Machine (SVM).

RT-Fall [11] is based on WiFall to improve the classification performance and to evaluate more motions. RT-Fall uses the amplitude and phase information of CSI for the signal segmentation and the feature extraction. The extracted features are based on those of WiFall and the phase information. The experimental results showed that this method classified more motions with higher accuracy than WiFall did.

However, WiFall and RT-Fall have a problem of degradation of the classification performance when the classifier uses the data in different environments for learning and testing. When a classifier that has been trained on data measured in one room is applied to data measured in another room, the classification performance degrades.

B. Time-Frequency-Domain-Based Method

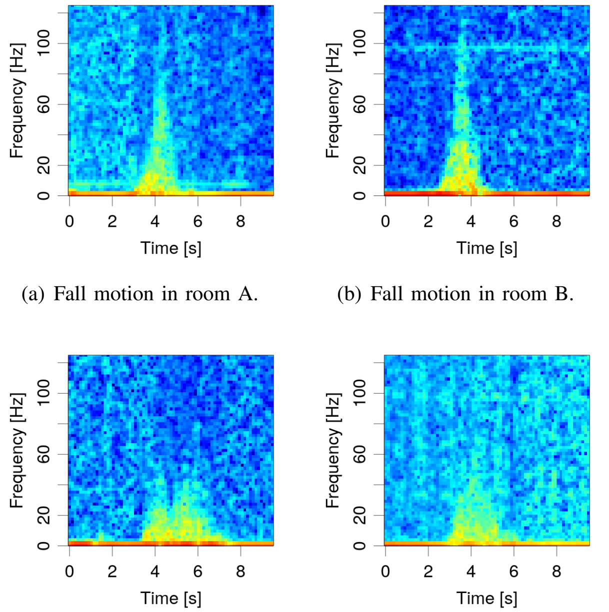
FallDeFi tried to tackle the performance degradation of WiFall and RT-Fall when the classifier uses the data in different environments for the learning and the testing [12]. FallDeFi is based on Time-Frequency domain information. For the selected CSI, a spectrogram is calculated, and the noise reduction of the spectrogram is applied to separate the motion part from the background noise, cluttering, and some outliers. The signal segmentation is performed to detect the fall candidate by comparing the background noise and the sum of spectral power between 5 Hz to 25 Hz, and features are extracted from time-frequency domain information of the segmented signal. Fall and other motions are measured in multiple environments and the classification performance is evaluated in multiple environments.

However, if the SNR of the signal is small, the classification performance degrades. This is because noise reduction during spectrogram processing removes weak energy part of the motion, which leads to the extraction of non-accurate features. In addition, there is a possibility that a candidate signal of a motion is extracted even if there is no motion.

IV. Proposed Method

A. Preliminary Experiment

To check the similarity of the CSI change due to a fall motion in different environments, we recorded such CSI signals in two different rooms. Fig. 1 shows the spectrograms of the motion in the two different rooms by two different people. There is no significant difference of the spectrograms in two different rooms. In Fig. 1(a) and 1(b), the spectrograms



(a) Fall motion in room A. (b) Fall motion in room B.

(c) Lying down motion in room A. (d) Lying down motion in room B.

Fig. 1. Spectrograms of the motion in two different rooms.

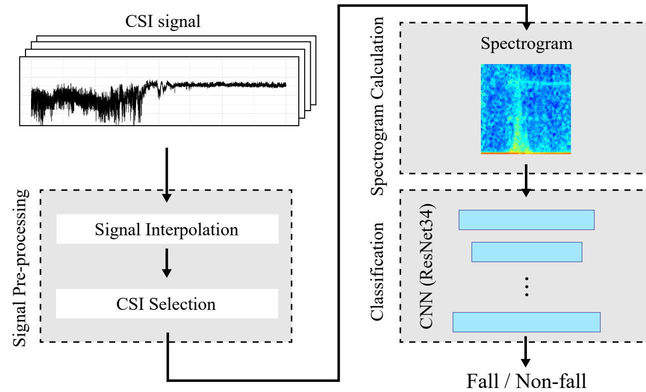


Fig. 2. A flowchart of the proposed method.

of the fall motion in both rooms are similar in terms of peak value, motion duration, and overall shape of the spectrogram. In Fig. 1(c) and 1(d), the spectrograms of the lying down motion are different from those of the fall motion in terms of peak value and motion duration.

B. Overview of Proposed Method

From the empirical study, we find that the spectrograms of same motions in two different rooms are not significantly different. Our proposed method is a spectrogram-image-based fall detection using CNN (Convolutional Neural Network). Unlike FallDeFi, CSI is segmented with the sliding time window, and then the fall is detected based on the spectrogram images from the segmented CSI each time window. Fig. 2 shows a flowchart of the proposed method. Our method consists of 3 steps: (i) signal pre-processing, (ii) spectrogram calculation, and (iii) classification.

C. Signal Pre-processing

There are two purposes of signal pre-processing: (i) make not-evenly-spaced packets evenly spaced, and (ii) select proper signals that reflects human motion. To achieve these purposes, interpolation and Principal Component Analysis (PCA) are done in this part.

1) Signal Interpolation: We collected CSI data by sending a ping between a transmitter and a receiver at certain time intervals (i.e. sampling rates). Because of the latency of packet arrival and/or the packet losses, the intervals of each CSI are not the same. To obtain the CSI data arranged at certain intervals, we apply linear interpolation to time series of the CSI data.

2) CSI Selection: Wang *et al.* found a correlation of signal variations for each antenna pair and subcarrier when a motion occurs from the experimental observation [10]. When a person is doing some motions, time variations of CSI over subcarriers in the particular antenna pair are mostly correlated. In contrast, time variations of CSI from the different antenna pairs sometimes have a low correlation. Moreover, using the whole CSI is very costly in terms of computation time. One way to solve this problem is to select the proper CSI from the entire set of CSIs. However, identifying which CSI reflects the human motion the most is not straightforward and depends on the environment. Therefore, we apply PCA to get the CSI that reflects human motion.

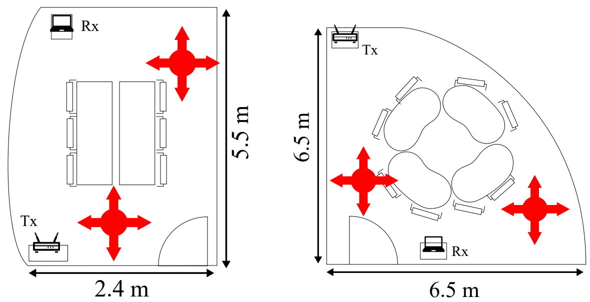
In PCA, the first principal component (PC) corresponds to the largest change part of CSI, which, basically, is supposed to include the human motion. We pick up some PCs to emphasize the human motions in the part of spectrogram calculation. To determine the number of PCs to be selected, we used cumulative proportion of variance. Proportion of variance of the ith principal component v_i is expressed as a ratio of eigenvalue of the ith principal component λ_i to the sum of the all eigenvalues, which is shown in Eq. (3).

$$v_i = \frac{\lambda_i}{\displaystyle\sum_{k=1}^{N_h} \lambda_k} \tag{3}$$

The ith cumulative proportion of variance is a sum of the proportion of variance from the first PC to the ith PC. We selected PCs from the first to which cumulative proportion of variance exceeds 0.90 firstly.

D. Spectrogram Calculation

After doing PCA, we generated a spectrogram image from the selected PCs. For each selected PC, we apply short time Fourier transform (STFT) with 512 ms window size and 128 ms overlap, and calculate spectrogram. In some conventional methods, for each activity, averaging spectrogram or signals is used to obtain a single one from CSI data [9], [12], [15]. To emphasize the motion part of the spectrogram, we obtain



(a) Room A (b) Room B

Fig. 3. Experimental environments.

a single spectrogram by using the weighted average based on proportion of the variance of PC. The weighted averaging from the first PC to the nth PC is expressed as $S = \sum_{k=1}^{n} v_i S_i$, where S is the weighted average of spectrogram, and S_i is the ith spectrogram.

After the spectrograms are averaged, the spectrogram image is generated from the averaged spectrogram. Since CNN uses RGB pixel values, the colormap is the same for all spectrograms to keep the color in which the spectrogram values are represented. In addition, from the experimental study, we found that falling motions do not contain the frequency response above 125 Hz. Thus, we cut the spectrogram above 125 Hz. Since the used CNN is pre-trained with ImageNet dataset which is 227×227 RGB images [16], we generate images with the same vertical and horizontal size and resized them to a size of 227×227.

E. Classification

In this section, we explain the structure of the classifier and the procedure of learning and classification. As we mentioned first, we use spectral-image-based fall detection using CNN. By using CNN, we aim to capture the characteristic of human motion which does not depend on an environment. We use ResNet34 [13] model, which consists of 34 convolutional layers. The used model is pre-trained with ImageNet dataset [16]. ImageNet is a large scale dataset which consists of 1.5 million RGB images. Although a spectrogram image is an artificial image which is different from ones in ImageNet dataset, it is more effective to start with pre-trained weights than to start with weights set randomly. This allows the convergence to be faster, and overcomes the issue of the small size of our data set. In the offline learning part, we train spectrogram images collected in a single room to the ResNet34 CNN, and Resnet34 is initially trained with the ImageNet dataset.

V. Experimental Evaluation

In this section, we first explain the experimental setup, and then show the experimental results.

A. Experiment Setup

1) Hardware and Software: We used consumer Wi-Fi devices to collect data. As a transmitter, we used Buffalo WZR-1750DHP router which runs on 5 GHz IEEE802.11n and 20 MHz-bandwidth. As a receiver, we used a laptop which is equipped with the IEEE 802.11n Intel WiFi Link 5300 chipset. The transmitter and the receiver communicate by 2x2 MIMO, therefore CSI data has 4 antenna pairs for each subcarrier. We used the CSI tool proposed by Halperin *et al.* [17]. Because of the limitation of the tool, the selected 30 subcarriers are captured for each antenna pair. Thus, the number of elements of CSI is 120.

TABLE I
Number of recorded motions.

Room	Motion Type	Amounts
A	Fall	147
	Non-fall	219
B	Fall	195
	Non-fall	223

2) Data collection: To evaluate the classification performance over the environment, we collected data in different rooms, which is depicted in Fig. 3. The volunteers performed some motions at red point, the arrows indicates the direction of fall motion. In the experiments the volunteers performed: (i) fall, (ii) sit on the chair, (iii) sit on the floor, (iv) stand up from the chair, (v) stand up from the floor, (vi) lie down, (vii) pick up, (viii) jump once, (ix) walking, and (x) upper body motion. We measured these motions within 10 second duration. The falling was measured in the form in which the volunteers imitated the action on the assumption of falling by swoon and falling by sliding. We used ping for collecting CSI, and we set the time interval of sending ping (sampling rate) to 1ms/s (1000 Hz). The number of measured motions is shown in Table I.

3) Classification Specification: We conducted binary classification to identify fall motion from the other non-fall motions. In the classification, the whole data from one room is used as the training set, the data from the other room is used as the testing set. In the first part of the online training, all the layers, except the last fully-connected one, are frozen. The training is done this way for 16 epochs. Afterwards, all the layers are unfrozen, and we continue the training for 32 more epochs. When training only the last fully-connected layer, we set the learning rate to 10^{-2}. This would allow faster convergence. Afterwards, in the second part where the entire network is trained, we set the learning rate cyclically between 10^{-2} and 10^{-6} according to [18]. The performance metrics used to evaluate our method are accuracy, precision, recall, and the F1 score.

B. Comparison of the CNN

To confirm the effectiveness of the pre-trained CNN of the proposed method, we first compared the classification performance between the different CNN architectures. We use two "shallow" CNN to compare the classification performance with ResNet34. The first one consists 3 convolutional layers and 3 dense layers. The second one consists 6 convolutional

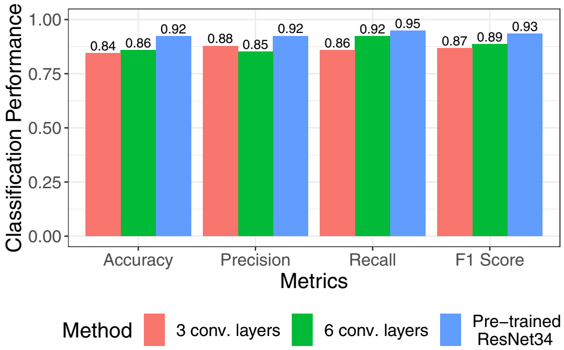
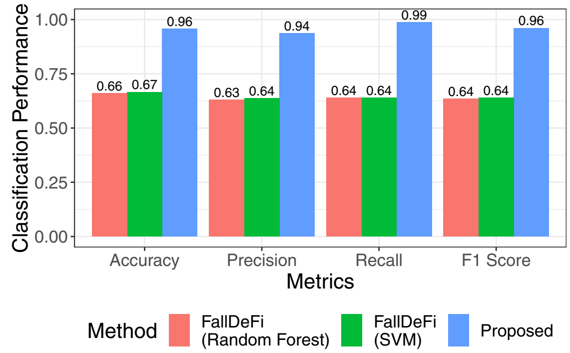


Fig. 4. The classification performance using different CNN is used for learning and classification.

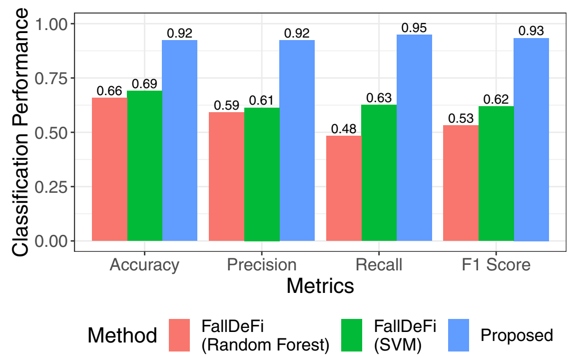
layers and 3 dense layers. Both CNNs are not pre-trained with the ImageNet dataset, and the number of layers in each is much smaller than that of ResNet34. The room B dataset is used for learning, and the room A dataset is used for testing. The classification performance of the different CNNs is shown in Fig. 4. Shallow Networks, with 3 convolutional layers and 6 convolutional layers reached accuracies equal to 0.84 and 0.86, respectively. This is clearly much lower than the accuracy obtained by using the pre-trained ResNet34 which is equal to 0.92. This trend is also observed among the other metrics as the pre-trained ResNet34 has the best performance. From the result, it is considered that the transfer learning for the fall detection task is effective for the CNN which is pre-trained with the ImageNet dataset. On the other hand, the two non-pre-trained CNNs suffered from overfitting, since the amount of data is not enough to train the parameters in the CNN. By using the pre-trained model, it was possible to avoid this problem. In other words, fitting a general model into our specific task might be more effective than training from scratch. As suggested before, this is mainly due to the limited amount of data in our hands.

C. Comparison with FallDeFi

We compared the classification performance of our proposed method with that of FallDeFi [12]. To extract the features of FallDeFi, we used the author's FallDeFi implementation available in [19]. As a classifier, we used SVM which is used in original paper [12], and we additionally used Random Forest. The features we used for classification are as follows: (i) event duration, (ii) spectral entropy (1 to 10 Hz), (iii) spectral entropy (10 to 30 Hz), (iv) fractal dimension, which marked the best classification performance according to the original paper [12]. The classification performance of our proposed method and FallDeFi are shown in Fig. 5. Fig. 5(a) shows the classification performance of the proposed method and FallDeFi when the data of the room A are used for learning and training and that of the room B are used for evaluation and testing. Fig. 5(b) shows the result with the room data used for learning and testing set reversed from that in Fig. 5(a). In our proposed method, both the results of two evaluations performed over 0.90 accuracy. In contrast,



(a) The results when Room A data are used for learning and Room B data are used for testing

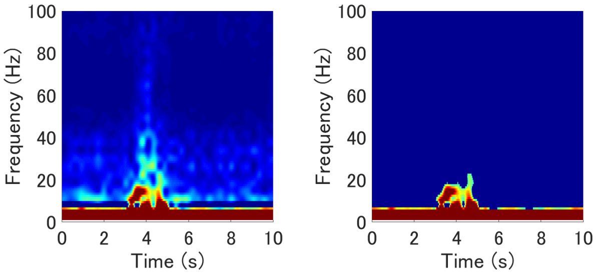


(b) The results when Room B data are used for learning and Room A data are used for testing

Fig. 5. The classification performance using different room data for learning and testing.

FallDeFi marked 0.65 accuracy for both SVM and Random Forest classifiers. Recall refers to the ratio of fall data that was correctly identified. If the fall event could not be detected, it results in a serious risk of human life. In our proposed method, the recall marked over 0.95, while it marked about 0.55 in the FallDeFi. That means FallDeFi missed about the half of fall motion. In the precision and the F1 score, our proposed method performed higher results than those of FallDeFi.

The reason of these performance is that the features of the motions were captured in detail in our proposed method. In FallDeFi, some image processing for noise reduction are applied to separate signal of the motion from some background noises and cluttering. During this process, the noises are replaced in zero, which can cause the distortion of the original signal. Fig. 6 shows the spectrograms from the FallDeFi implementation [19]. Fig. 6(a) and Fig. 6(b) show a spectrogram of fall motion before and after noise reduction, respectively. In Fig. 6(a), the spectrogram of motion can be confirmed in the range of 40 Hz or less around the point of 4 second. In Fig. 6(b), the shape of spectrogram is changed in the result of replacing weak energy part into zero. The distortion due to this process means that the features of a motion cannot be accurately extracted. Furthermore, since the threshold is used to segment the motion candidate, the signal may be missed if the motion signal is weak. If there is a low frequency noise



(a) Before noise removal. (b) After noise removal.

Fig. 6. Comparison of the spectrograms of the fall motion in FallDeFi process.

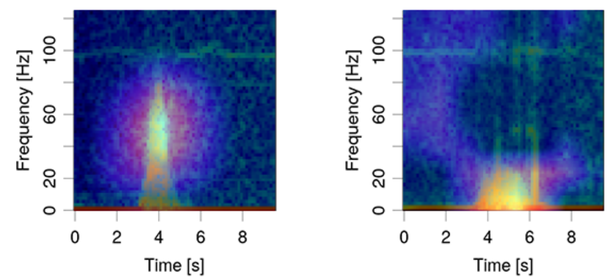
in the CSI, the fall candidate is segmented even when there is no motion, and thus features are not accurately extracted.

Fig. 7 shows the heatmap of the spectrogram images. The heatmap shows which part of the image is used for classification in the CNN of our proposed method. CNN uses the peak part of the fall motion in the spectrogram in Fig. 7(a), and the low frequency part of the non-fall motion is used in Fig. 7(b). Since CNN uses the different parts of the fall and the non-fall, CNN can capture the characteristic of motions which is common in the environment.

In our proposed method, the computational time is considered to be large because signal pre-processing is applied to 120 CSIs. If accidents should be detected immediately, the computational time will be the bottleneck when we cannot expect a computational power. However, in many situations, the system can have time for computation, for example, we can have some time to detect and inform an accident, such as falling.

VI. Conclusion

In this paper, we proposed a spectrogram image-based fall detection using Wi-Fi CSI (Channel State Information). Unlike the conventional method, CSI is segmented with the certain sliding time window, and then the CNN detects fall by using the spectrogram image from the segmented CSI. We carried out experiments to evaluate the classification performance in different two rooms and compared the classification performance with the conventional method. As a result, we confirmed that our proposed method outperformed conventional one and reached 0.90 accuracy.



(a) Fall motion. (b) Non fall motion.

Fig. 7. The CNN heatmap of spectrogram images. The highlighted part used for deciding class.

References

[1] W. H. O. Ageing, F. Life Course, and C. H. Unit, *WHO Global Report on Falls Prevention in Olser Age*. World Health Organization, 2008.

[2] M. Mubashir, L. Shao, and L. Seed, "A survey on fall detection: Principles and approaches," *Neurocomputing*, vol. 100, pp. 144–152, Jan. 2013.

[3] M. Cornacchia, K. Ozcan, Y. Zheng, and S. Velipasalar, "A Survey on Activity Detection and Classification Using Wearable Sensors," *IEEE Sensors J.*, vol. 17, no. 2, pp. 386–403, Jan. 2017.

[4] T. Liu, X. Guo, and G. Wang, "Elderly-falling detection using distributed direction-sensitive pyroelectric infrared sensor arrays," *Multidimens. Syst. Signal Process.*, vol. 23, no. 4, pp. 451–467, Dec. 2012.

[5] Z. Zhang, C. Conly, and V. Athitsos, "A survey on vision-based fall detection," in *Proc. 8th ACM Int. Conf. PErvasive Technol. Relat. to Assist. Environ.* ACM, 2015, pp. 46:1–46:7.

[6] N. Lu, Y. Wu, L. Feng, and J. Song, "Deep learning for fall detection: Three-dimensional CNN Combined with LSTM on video kinematic data," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 1, pp. 314–323, Jan. 2019.

[7] Z. Peng, J.-M. Munoz-Ferreras, R. Gomez-Garcia, and C. Li, "FMCW radar fall detection based on ISAR processing utilizing the properties of RCS, range, and Doppler," in *IEEE MTT-S Int. Microw. Symp.* IEEE, May 2016, pp. 1–3.

[8] S. Tao, M. Kudo, and H. Nonaka, "Privacy-Preserved Behavior Analysis and Fall Detection by an Infrared Ceiling Sensor Network," *Sensors*, vol. 12, no. 12, pp. 16 920–16 936, Dec. 2012.

[9] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and Modeling of WiFi Signal Based Human Activity Recognition," in *Proc. 21st Annu. Int. Conf. Mob. Comput. Netw.* ACM, 2015, pp. 65–76.

[10] Y. Wang, K. Wu, and L. M. Ni, "WiFall: Device-Free Fall Detection by Wireless Networks," *IEEE Trans. Mob. Comput.*, vol. 16, no. 2, pp. 581–594, 2017.

[11] H. Wang, D. Zhang, Y. Wang, J. Ma, Y. Wang, and S. Li, "RT-Fall: A Real-Time and Contactless Fall Detection System with Commodity WiFi Devices," *IEEE Trans. Mob. Comput.*, vol. 16, no. 2, pp. 511–526, Feb. 2017.

[12] S. Palipana, D. Rojas, P. Agrawal, and D. Pesch, "FallDeFi: Ubiquitous Fall Detection using Commodity Wi-Fi Devices," *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 1, no. 4, pp. 155:1–155:25, Jan. 2018.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv: 1512.03385*, 2015.

[14] N. Cummins, S. Amiriparian, G. Hagerer, A. Batliner, S. Steidl, and B. W. Schuller, "An image-based deep spectrum feature representation for the recognition of emotional speech," in *Proc. 25th ACM Int. Conf. Multimed.* ACM, Oct. 2017, pp. 478–484.

[15] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using wifi signals," in *Proc. 2016 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput.* ACM, 2016, pp. 363–373.

[16] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 248–255.

[17] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: gathering 802.11n traces with channel state information," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 1, p. 53, Jan. 2011.

[18] L. N. Smith, "Cyclical Learning Rates for Training Neural Networks," *arXiv: 1506.01186*, 2015.

[19] S. Palipana. (2017) Falldefi source code and data. [Online]. Available: https://github.com/dmsp123/FallDeFi (Accessed Oct. 16, 2019).