EasyChair Preprint

№ 6273

Wi-Fi-Based Human Activity Recognition Using Convolutional Neural Network

Muhammad Muaaz, Ali Chelli and Matthias Pätzold

August 10, 2021

Wi-Fi-Based Human Activity Recognition Using Convolutional Neural Network

Muhammad Muaaz[a], Ali Chelli[a], and Matthias Pätzold[a]

[a]Faculty of Engineering and Science, University of Agder, P.O. Box 509, 4898 Grimstad, Norway.

ARTICLE HISTORY
Compiled November 30, 2020

ABSTRACT
Unobtrusive human activity recognition plays an integral role in a lot of applications, such as active assisted living and health care for elderly and physically impaired people. Although existing Wi-Fi-based human activity recognition methods report good results, their performance is susceptible to changes in the environment. In this work, we present an approach to extract environment independent fingerprints of different human activities from the channel state information. First, we capture the channel state information by using the standard Wi-Fi network interface card. The channel state information is processed to reduce the noise and the impact of the phase offset. In addition, we apply the principal component analysis to removed redundant and correlated information. This step not only reduces the dimensions of the data but also removes the impact of the environment. Thereafter, we compute the spectrogram from the processed data which shows the environment independent fingerprint of the performed activity. We use these spectrogram images to train a convolutional neural network. Our approach is evaluated by using a human activity data set collected from 9 individuals while performing 4 activities (walking, falling, sitting, and picking up an object). The results show that our approach achieves an overall accuracy of 97.78%.

KEYWORDS
Activity recognition; convolutional neural network; principal component analysis; spectrogram

1. Introduction

Wi-Fi-based human activity recognition (HAR) has become an important research topic due to the growing number of applications that need to monitor indoor human activities in a truly unobtrusive way. These applications include elderly care, surveillance, and active assisted living. Furthermore, Wi-Fi-based HAR offers several advantages over vision- and wearable sensor-based techniques. For instance, in contrast to vision-based HAR systems, Wi-Fi-based systems are cost-effective, unaffected by lighting conditions, and preserve the user's privacy. Furthermore, the users are not required to wear the sensor in contrast to wearable sensor-based HAR systems.
In Wi-Fi-based HAR systems, a transmitter and a receiver are deployed in the environment. The transmitter emits radio signals, and the presence of moving objects in the propagation environment causes the Doppler frequency shift in these radio signals before they are received by the receiver. In the literature, it has been shown that the

CONTACT Muhammad Muaaz. Email: muhammad.muaaz@uia.no

received signal strength indicator (RSSI) and the channel state information (CSI) can be used to recognize human activities (Wang et al. 2017). In contrast to the RSSI which represents the attenuation of the received signal strength during propagation, the CSI is more informative. The CSI includes both amplitude and phase information associated with each orthogonal frequency division multiplexing (OFDM) subcarrier. It has been shown that CSI-based HAR systems generally perform better than RSSI-based HAR systems (Wang et al. 2017). There exist various approaches to recognize human activities from CSI data by using machine learning (Wang et al. 2017) and deep learning (Chen et al. 2018; Zou et al. 2018) techniques. In (Wang et al. 2017), the authors proposed two theoretical models. The first model (also known as "the CSI-speed model") links the speed of human body movements with the CSI data, while the second model (known as "the CSI-activity model") links the speed of human body movements with human activities (Wang et al. 2017). The proposed approach was developed using commercial Wi-Fi devices and achieved an overall recognition accuracy of 96%. In (Chen et al. 2018), an attention-based bidirectional long short-term memory (ABLSTM) technique was used to recognize humans activities from the CSI data. The CSI data sets collected in two different environments, namely an activity room and a meeting room were used to evaluate the performance of the proposed approach. This approach achieved recognition accuracies of 96.7% and 97.3% when the CSI data from the activity room and the meeting room is used, respectively. However, in the cross-environment scenario, where the training data that has been collected in one environment and the testing data from the other environment are used, the overall recognition accuracy drops to 32%. A deep learning technique consisting of autoencoder, convolutional neural network (CNN), and long short-term memory (LSTM) modules to recognize human activities from the CSI data has been proposed in (Zou et al. 2018). This deep learning network achieved an overall accuracy of 97.4%.

Although existing CSI-based HAR systems have reported reasonably good results, they still suffer from the drawback that they are environment dependent. This implies that their performance is susceptible to changes in the environment. One solution to this problem is to extract features from the CSI data that are subject and environment independent. This approach requires a lot of training data that must be collected from a variety of subjects in different environments (Jiang et al. 2018). The other approach proposes the use of a semi-supervised learning technique, which requires users to manually label the activity fingerprints that may have been changed due to changes in the environment (Wang et al. 2014). This solution requires user interaction that is not very practical for applications in elderly care.

In this work, we compute the spectrograms from the CSI data corresponding to different human activities. These spectrograms capture the Doppler characteristics of the radio channel caused by fixed and moving objects present in the environment. The static objects do not cause any variation in the trend of spectral components. This implies that different positions of static objects present in the environment will not influence the performance of our HAR system. These spectrograms are saved as portable network graphics (PNG) images and used to train a deep CNN. We evaluate this novel approach by using a CSI data set which is collected from 9 participants while performing four different activities: walking, falling, picking up an object, and sitting on a chair. Using this data set, our approach yields an overall accuracy of 97.78%. Moreover, our system recognizes the activities performed at greater distances. For instance, three out of the four activities are performed at a distance of 13 feet from the transmitting and receiving antennas.

The rest of the paper is organized as follows. Details about the experimental setup

and human activity data collection are given in Section 2. In Section 3, we explain the steps involved in processing the CSI data and computing the spectrogram. In Section 4, we present our CNN model, the classification process, and the obtained results. Finally, concluding remarks are given in Section 5.

2. Experimental Setup and CSI Data Collection

In this paper, we considered an indoor environment where 9 participants performed four different activities, namely walking, sitting on a chair, falling on a mattress, and picking up an object from the floor. During the data collection process, we ensured that only a single person is moving inside the room and all other objects are static. The participants of this experiment were asked to stand still for one second before starting an activity and after finishing that activity.

For the walking activity, we asked the participant to walk in a straight line from Point A to Point B and back (see Fig. 1). They repeated the activity 10 times, walking five times from Point A to B and five times from Point B to A. For the sitting activity, we placed a chair at Point B and asked the participants to stand still next to the chair facing the antennas and then sit on the chair as shown in Fig. 1. For the falling activity, a mattress was placed at Point B, and the participants were asked to stand on the shorter edge of the mattress and then fall on it. They repeated the activity 10 times. Out of these 10 falling trials, they fell on the mattress five times facing towards the antennas and five times facing away from the antennas. The last activity, picking up an object from the floor was also repeated five times, placing a small object on the floor at Point B, and asking the participants to pick it up.

To collect and parse the CSI data while the participants performed the activities mentioned above, we used two laptops, each was equipped with an Intel 5300 Wi-Fi network interface card (NIC). We installed the CSI Tool (Halperin et al. 2011) on both laptops. One laptop is used as the transmitter (T_x) and the other laptop as the receiver (R_x). The NICs of the T_x and the R_x are configured to operate at 5.745 GHz band with 20 MHz bandwidth in single-input multiple-output (SIMO) transmission mode. Instead of using the internal antennas of the laptops, which normally have a limited range, we connected an external directional antenna to the T_x and two external antennas to the R_x, where one of which was a directional and the other one

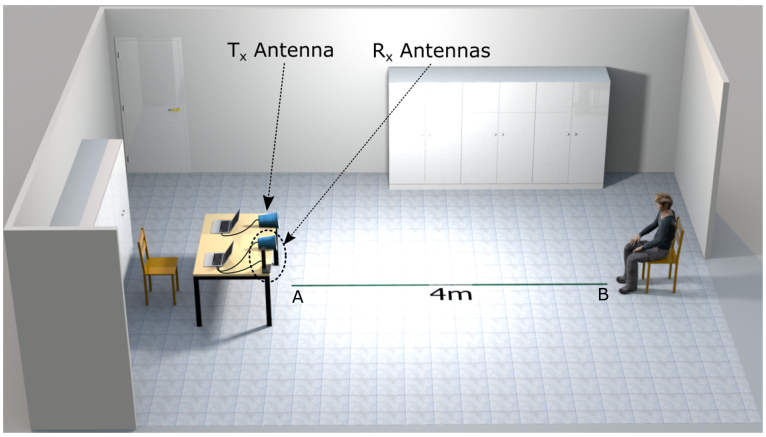

Figure 1. The experimental setup for data collection.

an omnidirectional antenna. We used the injector-monitor Wi-Fi mode, where the T_x was set to inject 1000 random data packets per second into the wireless channel, and the R_x captured the injected packets. The transmitting and receiving antennas were attached to a table as shown in Fig. 1 at a height of 0.8 meters from the floor. For each received data packet, the R_x reports the estimated CSI in a matrix form. The dimension of the CSI data matrix was $N_{T_x} \times N_{R_x} \times K$, where N_{T_x} indicates the number of transmit antennas, N_{R_x} stands for the number of receive antennas, and K represents the number of OFDM subcarriers. By default, the CSI tool reports estimated CSI data along 30 OFDM subcarriers for each transmission link. Therefore, in our case, the dimension of the CSI data matrix was $1 \times 2 \times 30$.

3. Processing CSI Data and Estimating the Spectrogram

The raw CSI data contains amplitude and phase information. Both the amplitude and phase of the CSI data are corrupted by noise; and therefore, the CSI data streams can not directly be used for activity recognition. The noise sources of the amplitude of the CSI data are mainly the ambient noise and adaptive changes of the transmission parameters (Yousefi et al. 2017). In addition to that, the phase of CSI data suffers from the carrier frequency offset (CFO) and the sampling frequency offset (SFO) (Wang et al. 2017; Yousefi et al. 2017). The errors related to the CFO and SFO are due to the asynchronicity between the transmitter and receiver clocks.

To denoise these data, we first calibrated the phase of the CSI data by applying the CSI ratio method (Zeng et al. 2019), where it has been shown that this approach significantly reduces the influence of CFO and SFO on the phase. The CSI ratio method requires that two antennas must be connected to the same receiver to collect the CSI data simultaneously. Thereafter, the CSI data from the first transmission link are divided by the CSI data of the second transmission link. Recall that each transmission link reports 30 CSI streams thus, we can obtain 30 CSI ratios. By comparing the spectrogram images of the CSI ratio method and the back-to-back phase calibration method (Keerativoranan et al. 2018), we observed in our experiments that the CSI ratio method works better.

Thereafter, we remove the correlated CSI ratios using the principal component analysis (PCA), which applies an orthogonal transformation to the 30 CSI ratios and converts them to 30 linearly uncorrelated variables. These variables are called principal components, where the first PCA component has the highest possible variance and the last PCA component the lowest variance. At this stage, we performed several experiments to determine the suitable number of PCA components for the subsequent steps. We observed that the first PCA component is sufficient to obtain a spectrogram that clearly shows the environment independent fingerprint of the performed activity as shown in Fig. 2.

To further minimize the effect of the high frequency components, which are not caused by the human movement, we apply a low pass filter to the selected principal component. Thereafter, we first compute the short-time Fourier transform (STFT) of the filtered data as given in (1). In (1), t', t, $y(t)$, and $g(t)$ indicate the running time, the local time, the filtered data, and the Gaussian sliding window function, respectively.

$$X(f,t) = \int_{-\infty}^{\infty} y(t')g(t'-t)e^{-j2\pi ft'}dt'. \tag{1}$$

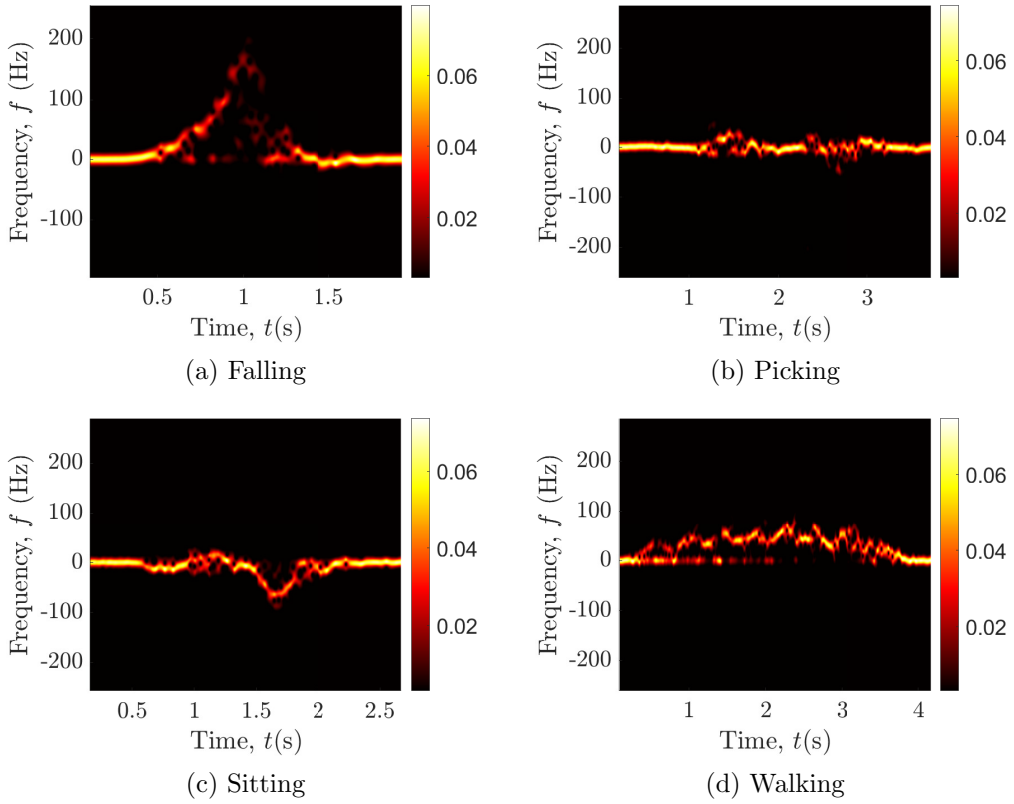

Figure 2. Spectrograms of the four activities.

Finally, the STFT ($X(f,t)$) is multiplied with its complex conjugate ($S(f,t) = |X(f,t)|^2$) which gives the spectrogram (Boashash 2015).

4. Classifying Spectrogram Images with CNN

For every activity trial in the collected data, we first computed the spectrogram and then saved it as a PNG image in a folder labelled with the activity. Thereafter, all spectrogram images were scaled to the same $224 \times 224 \times 3$ dimension by applying the bicubic interpolation technique. We split the spectrogram data into the train, validation, and test data sets representing 70%, 15%, and 15% of the total data. The training data were used to train the CNN (shown in Fig. 3) with a batch size of 16.

The CNN model consists of 14 layers including input, flatten, and output layers. The dimensions (i.e., height and width) of the filters used in all convolutional layers are 5×5 and in all max-pooling layers 2×2. The stride parameter (i.e., the number of cell shifts over the given data matrix) was set to 1 for the convolutional layers and to 2 for the max-pooling layers. The number of filters in the first, second, and third convolutional layer was 32, 48, and 64, respectively. All convolutional layers used the rectified linear unit (ReLU) activation function. After each max-pooling layer, a dropout layer (indicated by a green circle in Fig. 3) with a threshold 0.3 was used. The last two layers are fully connected (FC) with dimensions 256×1 and 84×1, respectively. The dimension of the output layer is 4×1 and uses the softmax activation function. The validation data were used to monitor the training progress of the CNN and to

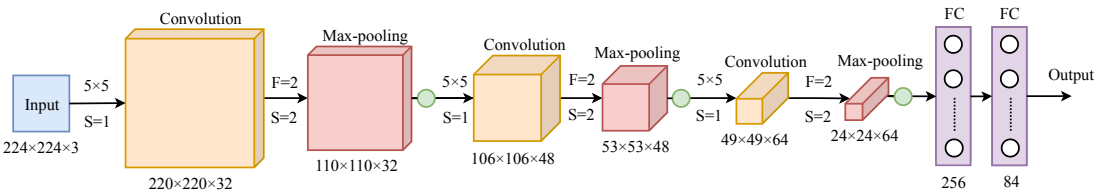

Figure 3. The architecture of the CNN, where the symbols S, F, and FC indicate stride, filter size of the max-pooling layer, and fully connected layer, respectively. The Green circles represent the dropout layer.

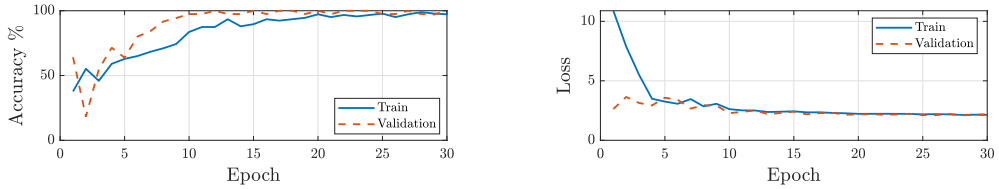

Figure 4. The accuracy (a) and loss (b) during the training process based on the training and validation data sets.

stop the training if the validation accuracy does not improve over 8 consecutive epochs. The accuracy and loss of the CNN model over the training and validation data are presented in Fig. 4(a) and Fig. 4(b), respectively.

Finally, the performance of the CNN model was evaluated based on the test data set. The results of our approach are shown in the confusion matrix (see Table 1). In this confusion matrix, the green cells represent the correctly classified examples and incorrectly classified examples are indicated in the red cells. The overall accuracy of the CNN model is given in the blue diagonal cell. We observe that the CNN model achieves an overall recognition accuracy of 97.78%. Moreover, the precision of the model for the activities walking, falling, picking up an object, and sitting is 100%, 93%, 100%, and 100%, respectively. The recall of the sitting activity is 88%, whereas the other three activities have a recall of 100%.

Table 1. The confusion matrix of results obtained from the CNN model.

-	-	Predicted labels				-
-	-	Walk	Fall	Pick	Sit	Precision
True labels	Walk	15	0	0	0	100%
	Fall	0	14	0	1	93%
	Pick	0	0	8	0	100%
	Sit	0	0	0	7	100%
-	Recall	100%	100%	100%	88%	97.78%

5. Conclusion

In this work, we developed a system that combines RF sensing and deep learning techniques to recognize human activities. In the RF sensing stage, we used two laptops, one acting as a transmitter and the other as a receiver to collect the CSI data. We collected CSI data while 9 participants performed 4 activities walking, falling, picking

up an object from the ground, and sitting on a chair. A three-step process was used to filter the collected CSI data. At first, we applied the CSI ratio method to the collected CSI data to reduce the impact of the phase offset. In the subsequent step, the PCA is applied to remove redundant and correlated information from the data. In the last step, a low pass filter is used to reduce the impact of high frequency components that were not caused by human movements. Thereafter, we computed a spectrogram for each activity trial in the collected data. These spectrogram images were divided into the train, validation, and test data sets. The training and validation data sets were used to train a 14-layer CNN model and monitor the training process, respectively. The test data set was used to evaluate the performance of the CNN model. The results show that our CNN model achieved an overall accuracy of 97.78%. In the future, we will conduct more experiments to quantitatively evaluate the performance of our approach in different environments.

Acknowledgement

This work is carried out within the scope of WiCare project funded by the Research Council of Norway under the grant number 261895/F20.

References

Boashash, Boualem. 2015. *Time-Frequency Signal Analysis and Processing – A Comprehensive Reference*. 2nd ed. Elsevier, Academic Press.

Chen, Zhenghua, Le Zhang, Chaoyang Jiang, Zhiguang Cao, and Wei Cui. 2018. "WiFi CSI-Based Passive Human Activity Recognition Using Attention Based BLSTM." *IEEE Transactions on Mobile Computing* 18 (11): 2714–2724.

Halperin, Daniel, Hu Wenjun, Sheth Anmol, and Wetherall David. 2011. "Tool Release: Gathering 802.11N Traces with Channel State Information." *ACM SIGCOMM Comput. Commun. Rev.* 41 (1): 53–53.

Jiang, Wenjun, et al. 2018. "Towards Environment Independent Device Free Human Activity Recognition." In *Proc. of the 24th Int. Conf. on Mobile Computing and Networking*, 289–304.

Keerativoranan, Nopphon, Haniz Azril, Kentaro Saito, and Takada Jun-ichi. 2018. "Mitigation of CSI Temporal Phase Rotation with B2B Calibration Method for Fine-Grained Motion Detection Analysis on Commodity Wi-Fi Devices." *Sensors* 18 (11).

Wang, Wei, X. Liu Alex, Shahzad Muhammad, Ling Kang, and Lu Sanglu. 2017. "Device-Free Human Activity Recognition Using Commercial WiFi Devices." *IEEE Journal on Selected Areas in Communications* 35 (5): 1118–1131.

Wang, Yan, et al. 2014. "E-eyes: Device-free Location-oriented Activity Identification Using Fine-grained WiFi Signatures." In *Proc. of the 20th Int. Conf. on Mobile Computing and Networking*, 617–628.

Yousefi, Siamak, Narui Hirokazu, Dayal Sankalp, Ermon Stefano, and Valaee Shahrokh. 2017. "A Survey on Behavior Recognition Using WiFi Channel State Information." *IEEE Communications Magazine* 55 (10): 98–104.

Zeng, Youwei, et al. 2019. "FarSense: Pushing the Range Limit of WiFi-Based Respiration Sensing with CSI Ratio of Two Antennas." *In Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3 (3).

Zou, Han, et al. 2018. "Deepsense: Device-free Human Activity Recognition via Autoencoder Long-term Recurrent Convolutional Network." In *2018 IEEE International Conference on Communications (ICC)*, 1–6.