High Accuracy Person Identification through Footstep-induced Vibration Signals

Ruipeng Deng[1], Xuebing Xu[2, *]
[1]Mathematics Department, Grinnell College, IA, USA
[2]School of Naval Architecture and Ocean Engineering, Huazhong University of Science and Technology, Wuhan, China
Email : dengruip@grinnell.edu , m201971730@hust.edu.cn

ABSTRACT: Person identification normally leverages biometrics which including face or voice recognition, fingerprint identification and so on. In these methods, people must actively participate in the activity for a successful identification. In addition, delicate software and hardware need to be applied for these methods. To overcome these restrictions, a novel person identification method is proposed by using people's footstep-induced vibrations. These time series vibration signals were first converted into two-dimensional time-frequency images by the continuous wavelet transform. Then deep convolutional neural network (CNN) models based on the Deep Residual Shrinkage Network (DRSN) were built to classify these images corresponding to each person respectively. Built on the test results of the CNN model, a new sensor layout strategy is developed which improved the identification success rate to 100%.

KEY WORDS: Person Identification, footstep-induced vibration signals, Deep Residual Shrinkage Network, Multi-sensor layout strategy.

• EXTENDED ABSTRACT

Non-intrusive pedestrian recognition is an important capability of smart buildings, which provides reference for the control system to plan the indoor environment and improve the security protection and living experience of users. Compared with existing recognition methods, the footstep ID based on vibration signals relies on a sparse sensor layout arranged in most scenes which provides an inexpensive and wide-coverage indoor perception for smart buildings. In addition, footsteps are not sensitive biometric features and can avoid unnecessary privacy disclosure. So, this research program aims to achieve a non-invasive person identification in smart buildings using footstep-induced vibration signals collected by geophone sensors placed on the pavement.

By analyzing the step events in the footstep signals, we proposed a pedestrian recognition method that combines the frequency domain analysis and convolutional neural network. The collected footstep signals are first converted into time-frequency images through the continuous wavelet transformation, which can reveal the effects of each pedestrian's unique walking pattern, the so-called gait, on the energy distribution of signals. Then, the deep residual shrinkage network was constructed with the residual shrinkage building unit as the core component. This identification model was trained and tested using data from each sensor respectively to provide us with recognition results from different locations. To further improve the reliability of pedestrian recognition, a new sensor layout strategy was proposed to combine the identification results from different sensors and reduce the misjudgments.

To validate the proposed method, an experiment containing footstep-induced vibration signals of different pedestrians was introduced to provide a dataset for indoor person identification. In the experiment, each pedestrian walked back and forth on the sidewalk at eight walking stride frequencies, and the resulting vibration signals were collected by sensors to identify five pedestrians. The experiment results show that the identification model can automatically extract the features related to pedestrian gait from the footstep-induced vibration signals, and achieves an average test accuracy of more than 95% on 5 sensors. Furthermore, by combining the results of three sensors, the final results of person recognition of 5 pedestrian were improved to 100%.

In conclusion, the proposed method can effectively capture information related to the gait in footstep-induced vibration signals, and combine the recognition results of multiple sensors to achieve recognition with higher accuracy and less misjudgments. The experiment results also demonstrate the broad prospects of footstep-induced vibration signals in non-intrusive pedestrian recognition, which provides a new and more flexible solution for indoor person identification in smart buildings.

1 INTRODUCTION

Person identification nowadays is increasingly critical to enhance safety in various applications. Current methods for identifying a person normally leverage the following three major mechanisms: Facial recognition, voice recognition and fingerprint-based identification. The facial recognition for security applications normally leverages machine learning techniques for the classification of different people [1]. However, under the uncontrolled scenario, the overall accuracy does not perform well. For voice recognition-based approach, researchers have used a novel high-speed pipelined architecture that supports real-time speaker recognition, detecting the authorized user with high accuracy [2]. But they tend to give unreliable results facing with the situation of identifying the voices of multiple individuals talking at the same time. The fingerprint-based authentication [3] uses the framework by means of fingerprint enhancement, feature extraction, and matching techniques. The minutiae-based feature extraction and matching approach suffers from critical issues, such as its computational complexity being usually very high. A biometric method [4] was proposed based on finger vein and face bimodal feature layer fusion, which uses a CNN and the fusion occurs in the feature layer. However, it is currently limited to computers with high GPU computing power and cannot be used on mobile devices, which greatly limits the application in practice.

In summary, current person identification methods by nature require sophisticated hardware such as camera or fingerprint reader. In addition, people have to actively engage into the data acquisition process such as pressing the fingerprints, looking at the camera or speaking in front of a microphone. Lastly, sufficient computing resources are necessary in order to guarantee the identification accuracy.

To overcome some of the restrictions in the above three methods, people have started to use footstep-induced vibrations for person identification. This method does not require people's active participation, and the vibrational signal can be collected while they are entering the building or walking on the floor at home. The data acquisition can be as simple as small-sized geophones [5, 6]placed at the corners of the room or hallway.

Current methods using footstep-induced vibration signals for person identification are as follows: Anchal [7] proposed using multiclass supervised machine learning algorithms such as SVM-Linear, SVM Gaussian, Logistic Regression, Linear Discriminant Analysis for person identification using footstep generated seismic signals. However, the accuracy is generally around 80% of the time which may not be the most optimal. Li [8] suggested an angle-constrained time difference of arrivals method to achieve identification using footstep-induced vibrations. However, the sensors used in the experiment are relatively large which can impede its real application. Pan [9] proposed using a system consisting of three modules: sensing, footstep analysis, and decision-making using footstep-induced structural vibration. One challenge is that the identification accuracy decreases as the number of users increases.

Figure 1.The problem stated in this study: using footstep-induced vibrations for person identification.

To overcome the above challenges and current limitations of using footstep-induced vibrations for person identification, in this study, we will propose a novel deep CNN based approach combined with continuous wavelet transform (CWT) for the preprocessing of the vibrational signals for high person accuracy identification. In particular, as shown in Figure 1, a person is walking on a floor which induces mechanical vibrations that will be collected by geophones. These vibration signals will then be the inputs to our model which then output the identity of the person.

2 METHOD OF APPROACH

The methodology in this study is shown in Figure 2. This method leverages a deep CNN model for person identification by integrating CNN with CWT. Built on the classification result from the model, a strategy was further developed to optimize the sensor layout.

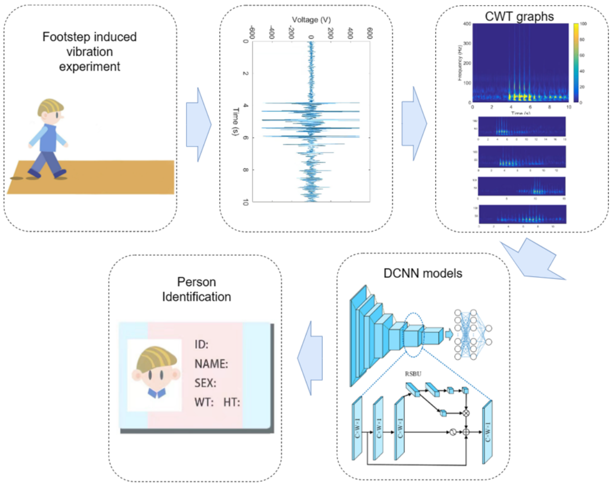

Figure 2. The flowchart of the proposed method.

2.1 *Signal Preprocessing Using CWT Based Approach*

The footstep-induced vibration signals are in the time domain, however, it also has the frequency domain characteristics. Therefore, performing signal analysis in either the time or frequency domain can be limited in terms of fully exposing the information stored in the signals. To overcome this challenge, this study uses the CWT which can be used to represent the energy variations within signals in both time and frequency domain. As shown in Figure 3, a typical time domain signal was converted into a 2D image by applying

CWT. This enriched information in the images can better help the later training process in terms of building a CNN model.

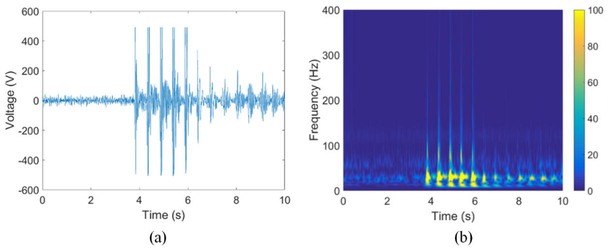

Figure 3. A sample of footstep-induced vibration signal. (a) The wave in time domain, (b) The CWT image in time and frequency Domain.

2.2 *CNN Model Applied in This Study*

CNN is a type of feedforward neural network which automatically extracts features through convolutional layers. As a specially designed CNN for signals with high noise, the Deep Residual Shrinkage Network (DRSN) model [10] introduces the soft thresholding into basic Residual Building Units (RBU) and possesses a new Residual Shrinkage Building Unit (RSBU). The RBU is one of the basic components of ResNet, whose input is x_l and output x_l can be expressed as:

$$x_{l+1} = h(x_l) + F(x_l, W_l) \tag{1}$$

where $h(x_l) = W_l x$ is a direct mapping generated through two convolutional layers and $F(x_l, W_l)$ is the residual part achieving by the identity shortcut. The identity mapping ensures that the network of layer l+1 must contain more image information than the previous layer, which reduces network redundancy. On this basis, the soft thresholding is introduced as a nonlinear layer to enhance the anti -noise ability of a new unit, namely RSBU. As shown below, soft thresholding is a segmentation function:

$$y = \begin{cases} x - \tau, \ x > \tau \\ \quad 0, \ -\tau \le x \ \le \tau \\ x + \tau, \ x < -\tau \end{cases} \tag{2}$$

where τ represents the threshold shrinking the input feature x towards zero and y represents the output feature. It is observed that the derivative of soft thresholding function is either zero or one, which is the same as ReLU to prevent gradient vanishing or gradient explosion during the training of model. In order to adaptively set thresholds for each channel, a small subnetwork is designed in RSBU.

As shown in Figure 4, threshold is determined by the mean of the absolute values of feature map and the coefficient α. Through the sigmoid function, α is limited to a number between 0 and 1, which effectively restricts the threshold to a positive number and not too large. Replacing the basic residual module in the original ResNet with RSBU, a structure of Deep Residual Shrinkage Network is obtained.

2.3 *Sensor layout optimization based on the sensor results*

After building the model on single sensor signals, a strategy is then proposed to further increase the identification accuracy by evaluating a combination of multiple sensors' performances.

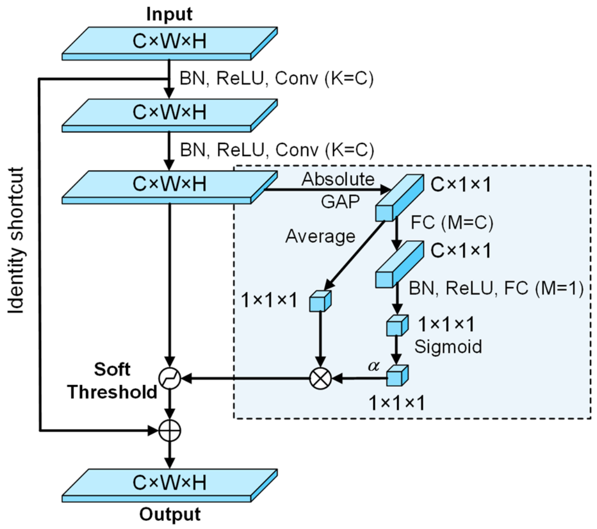

Figure 4. The architecture of the RSBU.

3 CASE STUDY

A case study was performed using the experiment results [11] from 5 different people by collecting footstep-induced vibration signals, which was to validate the proposed methodology.

3.1 *Introduction to the experiment*

This session will introduce briefly about the experiment from hardware, experiment setup and sample perspectives. The hardware applied in this experiment includes the key component Geophone associated with amplifier and processor [11]. As shown in Figure 5 which is the experiment setup, 5 sensors were placed at the corner of the hallway. Participants in this experiment walked through the hallway forward and backward during which the sensors recorded their footstep-induced vibration signals. The participants' walking stride frequency were also controlled by a metronome. 8 different walking stride frequencies were recorded for each participant. In summary, for each sensor, it will record all participants' data under various walking traces and walking stride frequencies.

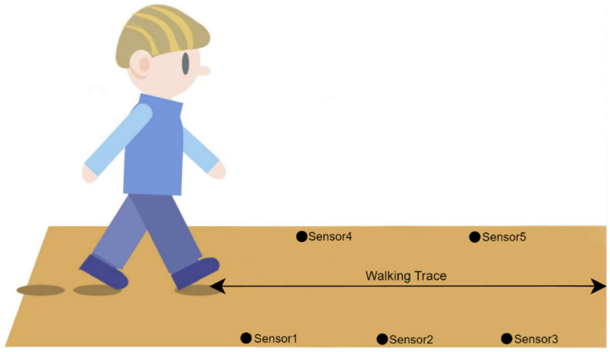

Figure 5. Experiment setup with sensor locations.

Table 1. Data collected in the experiment.

Person No.	1	2	3	4	5
No. of Signals Recorded	435	439	397	397	397

3.2 Preprocessing of the vibration signals using CWT

The experiment data applied in this study consist of a comprehensive number of footstep-induced signals for 5 different people from 5 sensors at different locations. The list of signals amount recorded for each person has been listed in Table 1. To expose the time-series signal of both time and frequency domain characteristics, CWT was used to convert these signals into 2D images. Unlike the analysis that focuses only on time or frequency domain, the CWT method is capable of describing the signal energy variations from both time and frequency scales, which provides enriched information to identify unique features conveyed in the signal. In addition, the CWT has multi-scale characteristics so that the signals can be examined from coarse to fine, enabling the demonstration of the local characteristics in signals from both the time and frequency domains. Thus, for the vibration signals, CWT is especially useful for the signal preprocessing [12].

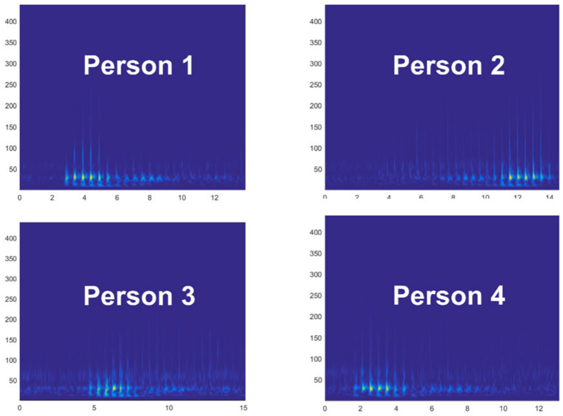

Figure 6. The CWT images of 4 pedestrians' footsteps collected by the same sensor.

The signals were first curated by normalization of their amplitude and converted to RGB images using Morse wavelet base in MATLAB. In Figure 6, we plotted the CWT images from 4 different people under the same speed from the same sensor. There is a perceivable difference on these images indicating the amount of energy represented in the signals varies. Another implication of this observation is that different people may have different intensity in their footsteps while walking.

3.3 Training, validating and testing of the CNN models

The next step is to construct a DRSN model and use the preprocessed CWT image dataset for training, validating, and testing. As shown in Table 2, this model includes 7 layers, and outputs a label value between 0 and 4 by inputting an image size of (224, 224, 3). More specifically, CONV_1 only involves one 3 × 3 convolutional layer for extracting preliminary features in the time-frequency images. CONV_2_X to CONV_5_X are comparatively more complex, with each layer encompassing two consecutively stacked RSBU modules. After successive convolution operations and dimension reduction by AVG_POOL, the features extracted from the input data are fed into FC to complete the final classification task.

In order to make the model with person recognition ability, this study constructs 5 different datasets, corresponding to the 5 sensors at different locations in the hallway shown in Figure 5. For example, dataset1 corresponding to sensor1 includes 5 categories (i.e. 5 people), total of 414 images. These images are split into training, validating and testing dataset in ratios of 7: 1: 2, respectively. Model 1 is obtained by alternately training and validating the DRSN model on the training and validating set, and is used predict the data collected by sensor 1. During the training of the model, the most important hyperparameter is the learning rate lr. The starting value of lr was set to be 0.001 and gradually decayed to 0 exponentially to make the decline process of the loss function more stable. Lastly, the performance of model1 was evaluated on the test set.

Table 2. The architecture of the person identification model.

Layer	Sequential	Output	Trainable parameters
CONV_1	Conv (64, 3, 1)	(224,224,64)	1,856
CONV_2_X	RSBU (64, 2, 1)	(224,224,64)	147,968
CONV_3_X	RSBU (128, 2, 2)	(112,112,128)	525,568
CONV_4_X	RSBU (256, 2,2)	(56,56,256)	2,099,712
CONV_5_X	RSBU (512, 2, 2)	(28,28,512)	8,393,728
AVG_POOL	BN, ReLU, GAP	(1,1,512)	0
FC	FC (5)	(5)	5130

4 RESULTS AND DISCUSSION

4.1 Test Results of the CNN model

This study built model 1 to 5 for person identification using footstep-induced signals from sensor 1 to 5 respectively. Take model 5 as an example, batch_size was set to 16 and alternately trained and verified for 100 times. After the completion of training process, the performance of the model was evaluated on the test set. As seen in the confusion matrix in Figure 7 (e), this model is able to achieve 100% recognition accuracy for person1, person 2 and person 5. However, for these 5 models in general, they possess a high confidence in identifying person1, person2, person 5 and misjudgments are concentrated in person 4. The summary of the training and testing results of each model are shown in Table 3. The test accuracy of all models surpasses 95%, of which model 2 has the highest accuracy of 98.75%.

Table 3. The average test accuracy of 5 models.

Model of sensors	Model 1	Model 2	Model 3	Model 4	Model 5
Testing accuracy (%)	97.50	98.75	96.25	97.50	95.00

4.2 Sensor Layout Optimization

Although the accuracy of every model is above 95%, it is observed that certain model performs better than others. This variation may be due to the sensor locations since each model is built on corresponding single sensor's data. Therefore, this motivates us to further optimize the accuracy and possibly reduce the number of sensors employed to achieve better results. Figure 8 demonstrates the proposed strategy by using multiple sensor model's results to make final identification decisions.

By observing the confusion matrix from all five models, it is noted that there can be redundancy in terms of using all five sensors for person identification. Therefore, we are proposing

Table 4. Improved identification accuracy using a combination of sensors.

Combination by person number	1,2,3	1,2,4	1,2,5	1,3,4	1,3,5	1,4,5	2,3,4	2,3,5	2,4,5	3,4,5
Identification accuracy (%)	**100**	98.75	98.75	98.75	98.75	**100**	98.75	98.75	97.5	97.5

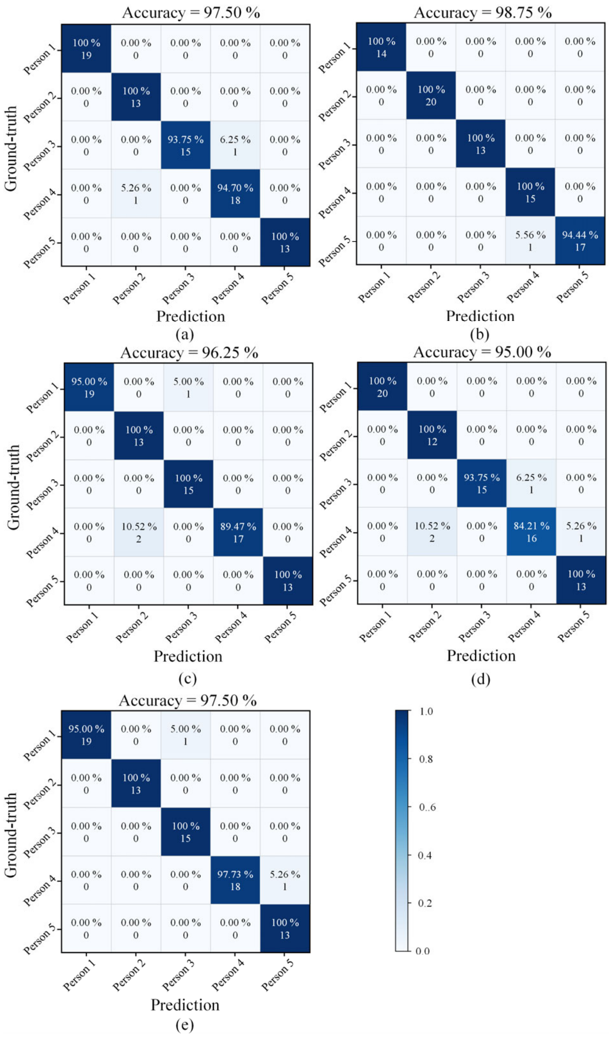

Figure 7. Confusion matrices of 5 models using footstep-induced signals from sensor 1 to 5: (a) model 1, (b) model 2, (c) model 3, (d) model 4, (e) model 5.

a method of using three sensors for the same purpose to further improve the identification accuracy other than using one sensor or all five sensors. The strategy applied here was the minority decision obeys the majority decision. If two sensors predict the same results, then the predication that the other sensor made would be ignored. For example, if two sensors made the right predication and the other sensor made the wrong decision, then the final predication that were made by those three sensors would still be right. On the other hand, if the two sensors give wrong predications, even if the third sensor made the right prediction, the result will still be wrong. By observing the predication result of five sensors, another rule was added. Person 4 was the hardest to be correctly identified. Sensor 2 was best in predicting person 4. Thus, whenever we are predicting person 4, only sensor 2 will make decision and ignore the other two sensors' prediction. To calculate the probability of using three sensors with above

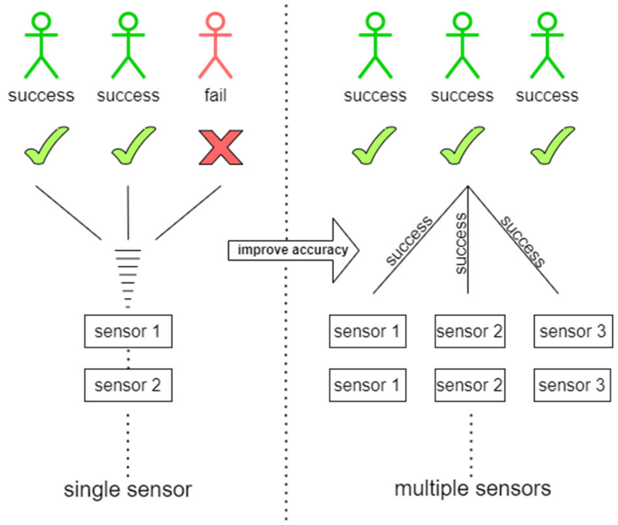

Figure 8. A schematic of the sensor layout optimization for improved identification accuracy.

strategies, R programming language was used. There are a total of 10 combinations of 3 sensors layout out of 5 sensors. After application of the strategy, the final predication rates were obtained and summarized in Table 4. As it is highlighted in Table 4, the combination of sensor 1, 2, 3 and sensor 1, 4, 5 both achieved 100% identification accuracy. This means that this strategy can further improve the accuracy. In the meantime, the results also indicates that it may not be necessary to use all five sensors in the experiment setup in order to achieve a decent identification accuracy.

5 CONCLUSION AND FUTURE WORK

This study applied a novel method by analyzing people's footstep-induced vibrations using deep CNN and CWT. The CWT was firstly applied to these time series vibration signals, which were converted to time–frequency images. Then deep CNN models based on DRSN were built for the classification of different people. Based on the model performances, a new strategy was proposed using a combination of sensors' results. By applying this strategy, the identification accuracy was further improved to 100%. Since current work involves 5 different participants in the experiment, future work can expand the number of participants for identifying this increased number of people. In addition, different textures of the floor will need to be tested because the floor materials can greatly influence the pattern of the collected vibration signals.

REFERENCES

[1] B. Chacua, I. Garcia, P. Rosero, L. Suarez, and M. Pusda, People Identification through Facial Recognition using Deep Learning, 2019 IEEE Latin American Conference on Computational Intelligence (LA-CCI). 2019.

[2] P. Dhakal, P. Damacharla, A. Y. Javaid, and V. Devabhaktuni, A Near Real-Time Automatic Speaker Recognition Architecture for Voice-Based User Interface, Machine Learning and Knowledge Extraction, 1, 1, 504-520, 2019.

[3] S. Bakheet, A. Al-Hamadi, and R. Youssef, A Fingerprint-Based Verification Framework Using Harris and SURF Feature Detection Algorithms, Applied Sciences, 12, 4, 2028, 2022.

[4] Y. Wang, D. Shi, and W. Zhou, Convolutional Neural Network Approach Based on Multimodal Biometric System with Fusion of Face and Finger Vein Features, Sensors, 22, 16, 6039, 2022.

[5] Y. Dong, J. J. Zou, J. Liu, J. Fagert, M. Mirshekari, L. Lowes, M. Iammarino, P. Zhang, and H. Y. Noh, MD-Vibe: physics-informed analysis of patient-induced structural vibration data for monitoring gait health in individuals with muscular dystrophy, Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers, Mexico, 525–531. 2020.

[6] Y. Dong, J. Fagert, P. Zhang, and H. Y. Noh, Non-parametric Bayesian Learning for Newcomer Detection using Footstep-Induced Floor Vibration: Poster Abstract, Proceedings of the 20th International Conference on Information Processing in Sensor Networks, 404–405. 2021.

[7] S. Anchal, B. Mukhopadhyay, and S. Kar, Person Identification and Imposter Detection Using Footstep Generated Seismic Signals, IEEE Transactions on Instrumentation and Measurement, 70, 1-11, 2021.

[8] F. Li, J. Clemente, M. Valero, Z. Tse, S. Li, and W. Song, Smart Home Monitoring System via Footstep-Induced Vibrations, IEEE Systems Journal, 14, 3, 3383-3389, 2020.

[9] S. Pan, N. Wang, Y. Qian, I. Velibeyoglu, H. Y. Noh, and P. Zhang, Indoor Person Identification through Footstep Induced Structural Vibration, Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications, 81-86. 2015.

[10] M. Zhao, S. Zhong, X. Fu, B. Tang, and M. Pecht, Deep Residual Shrinkage Networks for Fault Diagnosis, IEEE Transactions on Industrial Informatics, 16, 7, 4681-4690, 2020.

[11] Y. Dong, S. Pan, T. Yu, M. Mirshekari, J. Fagert, A. Bonde, O. J. Mengshoel, P. Zhang, and H. Y. Noh, The FootprintID Dataset: Footstep-Induced Structural Vibration Data for Indoor Person Identification with Different Walking Speeds, Zenodo, 2017.

[12] K. Zheng, Z. Li, Z. Ma, J. Chen, J. Zhou, and X. Su, Damage detection method based on Lamb waves for stiffened composite panels, Composite Structures, 225, 111137, 2019.