Device-free Surveillance using Non-imaging Sensors in a Sparsely Populated Outdoor Environment

Synopsis of the Thesis to be submitted in Partial Fulfillment
of the Requirements for the Award of the Degree of

Doctor of Philosophy

by

Priyankar Choudhary

Under the supervision of

Dr. Neeraj Goel and Dr. Mukesh Saini

Computer Science and Engineering

Indian Institute of Technology Ropar

October 28, 2022

1 Introduction

The proliferation and usage of microelectromechanical sensors [1] have made computing operations ubiquitous, mobile, and resilient. These sensors offer accurate and reliable statistics about the surrounding. As a result, numerous smart applications have been using sensors to monitor ongoing activities and behaviour. Further, these applications are employed to prevent, detect, and prepare a response to an event for safe and secure surroundings. Consequently, sensors are an indispensable component of security, surveillance, and reconnaissance applications.

Target information estimation using incoming sensor data streams demonstrates its applicability in numerous human-centric applications. Typically, in a human-centric application, first, a target (living or non-living) is detected. Then its location, identity, and activity-related information are estimated. Current applications require attaching a device such as GPS [2], smartphones [3, 4], wearable inertial sensors [5, 6], etc., to the body of the target to gather information. Further, these information are sent over the network to a storage and processing facility.

Owing to good accuracy and reliability, wearable sensor-based applications are very popular in many domains, such as medical [7] and entertainment [8]. However, there exists an intrinsic challenge, i.e, need to attach a sensing device to the target body. In some applications, it may not always be feasible to attach a device to the target, such as, wildlife monitoring, emergency rescue operations, surveillance, etc. Consequently, there is an increase in attention towards the usage of device-free (passive) sensing techniques. This alternative sensing strategy is attractive for many smart applications because a target does not need to be equipped with a sensing device. Moreover, there are no privacy violations or wear-and-carry inconveniences associated. Device-free techniques investigate and quantify the influence of a target on its surroundings. A few popular applications include logistics, health care, manufacturing and production, asset monitoring, security and defense, extraterrestrial body analysis, etc. [9].

Many device-free sensing techniques such as video [10], audio [11], passive infrared (PIR) [12], Bluetooth [13], ZigBee [14], and WiFi [15] are documented in the literature. Each of the aforementioned sensing techniques has a few drawbacks. Multipath propagation adversely affects the operation of WiFi, ZigBee, and Bluetooth signal-based applications, where different versions of the same signal reach the receiver at different times. The accuracy of PIR sensor-based applications can be compromised due to temperature fluctuations and no change in infrared radiation due to a static target [16, 17]. Video-based applications face the issues of occlusion, illumination, and privacy violations. Furthermore, video-based applications necessitate abundant storage, processing, and computing power [18].

On the other hand, audio and seismic sensors are device-free, non-invasive, inexpensive, and easy to install [18]. They are immune to temperature variations, occlusion, illumination [18], and electromagnetic disturbances [19]. Seismic sensors sense the earth's vibrations generated due to the movement of a target whereas the audio sensors are sensitive to surrounding noise, given the fact that human activities generate distinctive sound events. Hence, by detecting and classifying sound events can identify human activities in the vicinity. Additionally, seismic sensors are appropriate for both indoor and outdoor applications due to their low storage, processing, and computation power requirements.

2 Research Aim and Objectives

As surveillance applications, this thesis only focuses on localization and activity recognition of a single human target. The majority of previous research in seismic sensor-based localization is focused on developing applications for indoor scenarios. Indoor environments are less affected by changing environmental conditions over time. We intend to develop localization frameworks with use cases in outdoor scenarios. Moreover, we also focus on human-activity recognition as it is an important part of any surveillance application. The research aim and objectives of the thesis are summarized as follows:

1. First, we aim to identify different non-invasive and non-imaging device-free sensing techniques. Then, we choose the best available solutions suited for our experimental setup. Further, we conducted a comprehensive survey on available sensing techniques.

2. We aim to develop a module to estimate the location of the footsteps of a single human target. Real-world streaming data is challenging due to time-varying environmental conditions. So, it is important to develop a localization module that is immune to changes in the surrounding environment. We further explore whether the fusion of audio and seismic modalities could improve localization accuracy.

3. We aim to infer different human activities using a variety of sensing modalities (audio and seismic) and assess their efficacy.

4. Further, we aim to fuse the incoming data streams from different sensing modalities to improve the activity recognition performance over individual sensing modalities. The set of activities used for experiments is low-intensity sound and vibration generating activities.

3 Contributions of the Thesis

A detailed discussion on contribution of thesis is as follows:

3.1 A Survey on Seismic Sensor based Target Detection, Localization, Target Identification, and Activity Recognition

First, we provide a comprehensive overview of seismic sensor-based device-free surveillance. We have classified the survey into three categories, viz., (i) target detection, (ii) target localization, and (iii) target identification, and activity recognition. In this survey, we have covered how seismic sensors can be used as an alternative to video-based surveillance (i.e., CCTV cameras). Moreover, we have provided a brief overview of the seismic sensor along with its inherent challenges. It also explains the general architecture of seismic sensor-based applications.

3.2 Human Target Localization

After conducting a detailed survey, we focus on human target localization. The localization work is divided into two categories. In the first category, we only focus on developing unimodal localization frameworks using seismic sensors. We focus on developing a multimodal framework to localize a human target in the second category. The multimodal localization framework mitigates the limitations of unimodal frameworks. A detailed description of each category is as follows:

3.2.1 Seismic Sensor based Human Target Localization

In this work, we localize a human target by analyzing the trend of signal attenuation. The energy captured by a seismic sensor due to the movement of a target reduces with the increase in distance between the sensor and the source of vibration. A seismic sensor is omnidirectional in nature, so the sensing range around the sensor is considered to have a circular shape. The intersection point of three different circles with respect to three different sensors can be an estimated target location. To estimate the parameters of a circle, we use either a regression model or energy-distance relationship of seismic waves. Further, we found that circles may or may not intersect. So, we propose a heuristic to localize a human target. We developed a prototype system to evaluate the efficacy of the proposed methods. The localization modules show an average localization error of 1.37 meters in an area of 20.25 meter^2.

3.2.2 Audio-Seismic Fusion for Human Target Localization

In Section 3.2.1, a heuristic was proposed to localize a target. This work uses audio direction information as an alternative to the heuristic. The approaches fall into two groups. The first group of approaches applies a late fusion to the output of audio and seismic modalities for localization. First, a regression algorithm on the seismic signals is applied to estimate the distance of the target from each sensor; then, the target distance information is combined with the audio direction to infer the target location.

In contrast to the first group of approaches, the second group applies an early fusion of audio and seismic modalities to predict the target location directly. On different evaluation measures, we compare the proposed methods with multiple state-of-the-art. On 5-fold cross-validation, we achieve a root-mean-localization error of 0.735 and 0.907 meters in an area of 324 meter^2 for dataset-1 and dataset-2, respectively. For datasets 1 and 2, the best results with the proposed methods show an improvement of 4.68 and 4.57 meters over the best-performing state-of-the-art methods.

3.3 Human Activity Recognition

Similar to localization work, here also, we first focus on unimodal human activity recognition frameworks, followed by multimodal frameworks. The activity recognition work is divided into three categories. In the first and second categories, we have used seismic and audio modalities, respectively. We focus on developing a multimodal framework to recognize human activities in the third category. A detailed description of each category is as follows:

3.3.1 Seismic Sensor based Human Activity Recognition

In this work, we demonstrate the use of a seismic sensor for human activity recognition. Traditionally, researchers have relied on handcrafted features to identify the target activities, but these features may be inefficient in complex and noisy environments. The proposed framework employs an autoencoder to map the activity into a compact representative descriptor. Further, an artificial neural network classifier is trained on the extracted descriptors. We compare the proposed framework with multiple machine learning classifiers and a state-of-the-art framework on different evaluation metrics. On 5-fold cross-validation, the proposed approach outperforms the state-of-the-art in terms of precision and recall by an average of 10.68 and 23.36%, respectively. Detailed experimental analysis shows that a duration of 1.5–2.0 seconds is sufficient to recognize the targeted activities. The targeted activities include running, jogging, walking, jumping jacks, jumping, and inactivity. Detailed experimental analysis shows that a duration of 1.5–2.0 seconds is sufficient to identify the activity class. The targeted activities include run, jog, walk, jumping jacks, jump, and no activity.

3.3.2 Audio based Human Activity Recognition

There has been negligible attention to the recognition of low-intensity human activities for outdoor scenarios. In this work, we use a deep learning-based framework for recognizing different low-intensity human activities in a sparsely populated outdoor environment using audio. The targeted activities include running, jogging, walking, jumping jacks, jumping, hammer strikes on the ground, cycling, riding a bike, and no activity. The proposed framework classifies 2.0-second long audio recordings into one of 9 different activity classes. A variety of audio sounds in an outdoor environment makes it challenging to distinguish human activities from other background sounds. The proposed framework is an end-to-end architecture that employs a combination of mel-frequency cepstral coefficients and a 2D convolutional neural network to obtain a deep representation of activities and classify them. The extensive experimental analysis demonstrates that the proposed framework outperforms existing frameworks by 16.43% on the parameter F1-score.

3.3.3 Audio-Seismic Fusion for Human Activity Recognition

Activity recognition based on a single modality suffers due to some intrinsic challenges. The fusion of multiple sensing modalities can complement each other and thus produce the more robust and reliable results. In this work, we introduce a non-invasive human activity recognition system that utilizes footstep-induced vibration and sound in an outdoor environment with the aim of achieving improved performance over a single source of information. We employ 1D convolutional neural networks for automated feature extraction, fusion, and activity recognition on a 9-class classification problem. The proposed framework reports an average F1-score of 92%, which corresponds to a 5.74% improvement over the best-performing state-of-the-art. Confusion matrix-based analysis demonstrates that audio-seismic fusion not only reduces misclassifications but also reduces the impact of background noise on model performance. In addition, we demonstrate that a model trained on a balanced dataset has a higher F1-score than one trained on an imbalanced dataset. Activity-wise performance is reported to show the efficacy of the proposed fusion-based framework.

3.4 To collect and contribute suitable datasets for evaluation

Using a public dataset enables researchers to generate and compare various frameworks with the same input data. Unfortunately, there are no publicly available datasets for seismic sensors-based applications. It also makes it challenging to replicate the experiments in existing works. We collect and contribute datasets for evaluation and benchmarking purposes to the research community.

4 Thesis Organization

Chapter-wise thesis organization is summarized as follows:

Chapter 1 introduces the terms "device-free sensing" and describes key sensing technologies along with associated inherent limitations. It also discusses the motivation and challenges of opting for device-free sensing. We identify appropriate device-free sensing techniques and state the rationale behind this decision. Further, research gaps are also discussed.

Chapter 2 provides a comprehensive overview of the seismic sensor-based device-free sensing process and highlights the key techniques within the research field. We classify the existing literature into three categories, viz., (i) target detection, (ii) target localization, and (iii) target identification, and activity recognition. The techniques in each category are divided into multiple subcategories in a structured manner to comprehensively discuss the details.

Chapter 3 focuses on developing a working prototype of human target localization. We exploit attenuation in energy values with distance to localize a human target. We mathematically model the energy-distance relationship to learn circle parameters required for target localization. Furthermore, we have also used regression analysis to automatically learn the aforementioned relationship. A heuristic is also proposed to estimate the target location.

Chapter 4 discusses the effect of using multi-modal sensing modalities for target localization. In this work, we fuse audio direction information with seismic data to improve the localization accuracy that can be achieved using only seismic data.

Chapter 5 discusses the recognition of non-overlapping human activities using seismic data for single target scenarios. The proposed approach learns a deep representation of 16 distinct time and frequency domain features with reduced dimensionality using an autoencoder network. Further, the deep representation is given as an input to an artificial neural network for activity recognition. This chapter also investigates the optimal window length required for recognizing targeted activities using seismic data.

Chapter 6 examines the performance of the audio modality using 9 different human activities. The activity set used in this chapter is a superset of the activities used in Chapter 5. The proposed approach uses a combination of MFCC features and a 2D convolution neural network for recognizing target activities.

Chapter 7 uses audio-seismic fusion for human activity recognition. The proposed approach learns the deep representation of seismic and audio modalities utilizing a pair of 1D convolution neural networks. Further, the deep representation is given as input to a fusion block with the aim of achieving higher recognition accuracy than the use of a single sensing modality.

Chapter 8 concludes the thesis' contributions with future research directions.

5 Conclusions

The thesis systematically investigated how non-intrusive device-free sensing techniques are used to monitor a sparsely populated outdoor environment through six concrete research contributions. We first discussed various device-free sensing techniques before deciding on audio and seismic sensing modalities based on our surveillance requirements. Then, we surveyed state-of-the-art seismic sensor-based works focusing on detection, localization, target identification, and activity recognition. We start by localization of a human target using only a single sensing modality, i.e., seismic. We observed that the usage of seismic sensor alone show limited localization accuracy. Then, we decided to employ multiple sensing modalities because multi-modal information serves as a complementary source of information for each other, resulting in higher accuracy. Our proposed approach also demonstrated a higher localization accuracy on audio-seismic fusion.

Apart from the target location, recognizing its activity is also essential for a surveillance application. Similar to localization experiments, we first evaluated the performance of distinct sensing modalities (i.e., audio and seismic). We observed that individual sensing modality is affected by environmental noises, resulting in higher misclassification. Moreover, the nature of environmental noise is different for both audio and seismic sensing modalities. We also concluded that noises affect the audio modality more than the seismic modality. When we fused information from two different sensing modalities, we observed that the effect of background noise was significantly reduced, and system performance also increased. We also conducted experiments to estimate the effective duration of sensor readings to estimate the target activity. We found that a duration of 1.5-2.0 seconds is sufficient for target activity estimation.

6 Publications

Journals

- **Choudhary P**, Goel N, Saini M. A Survey on Seismic Sensor based Target Detection, Localization, Identification, and Activity Recognition, ACM Computing Surveys. 2022 Oct. (Accepted).

- **Choudhary P**, Kumari P, Goel N, Saini M. An Audio-Seismic Fusion Framework for Human Activity Recognition in an Outdoor Environment. IEEE Sensors Journal. 2022 Sep. 27;22(23):22817-27.

- **Choudhary P**, Goel N, Saini M. A Fingerprinting based Audio-Seismic Systems for Human Target Localization in an Outdoor Environment using Regression. IEEE Sensors Journal. 2022 Feb 25;22(8):7944-60.

Conferences

- **Choudhary P.**, Kumari P., Goel N. and Saini M. Low-Intensity Human Activity Recognition Framework using Audio Data in an Outdoor Environment. In proceedings of 7^{th} International Conference on Computer Vision and Image Processing 2023. Springer, Cham (accepted on September 30, 2022).

- **Choudhary P**, Goel N, Saini M. A Seismic Sensor based Human Activity Recognition

Framework using Deep Learning. In proceedings of 17^{th} IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) 2021 Nov. 16 (pp. 1-8).

- **Choudhary P**, Goel N, Saini M. Event Detection and Localization for Sparsely Populated Outdoor Environment Using Seismic Sensor. In proceedings of IEEE 6^{th} International Conference on Multimedia Big Data (BigMM) 2020 Sep. 24 (pp. 346-350).

References

[1] Chun Zhu and Weihua Sheng. "Wearable sensor-based hand gesture and daily activity recognition for robot-assisted living". In: *IEEE Trans. Syst. Man Cybern.: Syst.* 41.3 (Jan. 2011), pp. 569–573.

[2] Eleftherios Lygouras and Antonios Gasteratos. "A novel moving-base RTK-GPS-Based wearable apparatus for precise localization of humans in peril". In: *Microprocess. Microsyst.* 82 (Apr. 2021), p. 103833.

[3] Girija Chetty, Matthew White, and Farnaz Akther. "Smart phone based data mining for human activity recognition". In: *Procedia Computer Science* 46 (2015), pp. 1181–1187.

[4] Matthew Keally, Gang Zhou, Guoliang Xing, Jianxin Wu, and Andrew Pyles. "Pbn: towards practical activity recognition using smartphone-based body sensor networks". In: *Proceedings of 9^{th} ACM Conference on Embedded Networked Sensor Systems*. Seattle, WA, USA: ACM, 2011, pp. 246–259.

[5] Jennifer R Kwapisz, Gary M Weiss, and Samuel A Moore. "Activity recognition using cell phone accelerometers". In: *ACM SigKDD Explorations Newsletter* 12.2 (2010), pp. 74–82.

[6] Pierluigi Casale, Oriol Pujol, and Petia Radeva. "Human activity recognition from accelerometer data using a wearable device". In: *Proceedings of Iberian conference on pattern recognition and image analysis*. Berlin, Heidelberg, Germany: Springer, 2011, pp. 289–296.

[7] Arisa Olivia Putri, MA Ali, Mohammad Saad, and Sidiq Samsul Hidayat. "Wearable sensor and internet of things technology for better medical science: A review". In: *Int. J. Eng. Technol.* 7.4.11 (2018), pp. 1–4.

[8] JS Olson and S Redkar. "A survey of wearable sensor networks in health and entertainment". In: *MOJ Appl. Bionics Biomech.* 2.5 (2018), pp. 280–287.

[9] Stijn Denis, Rafael Berkvens, and Maarten Weyn. "A survey on detection, tracking and identification in radio frequency-based device-free localization". In: *Sensors* 19.23 (2019), p. 5329.

[10] Jason J Liu, Cody Phillips, and Kostas Daniilidis. "Video-based localization without 3D mapping for the visually impaired". In: *Proceedings of Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*. San Francisco, California, USA: IEEE, 2010, pp. 23–30.

[11] Xuehai Bian, Gregory D Abowd, and James Matthew Rehg. *Using sound source localization to monitor and infer activities in the Home*. Tech. rep. Georgia Institute of Technology, 2004.

[12] Suk Lee, Kyoung Nam Ha, and Kyung Chang Lee. "A pyroelectric infrared sensor-based indoor location-aware system for the smart home". In: *IEEE Trans. Consum. Electron.* 52.4 (2006), pp. 1311–1317.

[13] Vicente Canton Paterna, Anna Calveras Auge, Josep Paradells Aspas, and Maria Alejandra Perez Bullones. "A bluetooth low energy indoor positioning system with channel diversity, weighted trilateration and kalman filtering". In: *Sensors* 17.12 (2017), p. 2927.

[14] Boon-Giin Lee, Seung-Chul Lee, Wan-Young Chung, et al. "Enhanced RSSI-based high accuracy real-time user location tracking system for indoor and outdoor environments". In: *Int. J. Smart Sens. Intell. Syst.* 1.2 (2008), pp. 534–548.

[15] Moustafa Seifeldin, Ahmed Saeed, Ahmed E Kosba, Amr El-Keyi, and Moustafa Youssef. "Nuzzer: A large-scale device-free passive localization system for wireless environments". In: *IEEE Trans. Mob. Comput.* 12.7 (2012), pp. 1321–1334.

[16] Sang Gi Hong, Nae Soo Kim, and Whan Woo Kim. "Reduction of false alarm signals for PIR sensor in realistic outdoor surveillance". In: *ETRI J.* 35.1 (Feb 2013), pp. 80–88.

[17] J Andrews, A Vakil, and J Li. "Biometric Authentication and Stationary Detection of Human Subjects by Deep Learning of Passive Infrared (PIR) Sensor Data". In: *Proceedings of IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*. Philadelphia, PA, USA: IEEE, 2020, pp. 1–6. DOI: 10.1109/SPMB50085.2020.9353613.

[18] Satish G Iyengar, Pramod K Varshney, and Thyagaraju Damarla. "On the detection of footsteps based on acoustic and seismic sensing". In: *Proceedings of 41st Asilomar Conference on Signals, Systems and Computers*. Pacific Grove, CA, USA: IEEE, 2007, pp. 2248–2252.

[19] Meng Wang, Li Min, Xiaolei Zhang, Faxiang Zhang, Zhihui Sun, Shujuan Li, Chang Wang, Zhong Zhao, and Guanghu Hao. "Field test investigation of high sensitivity fiber optic seismic geophone". In: *Fiber Optic Sensing and Optical Communications* 10464 (2017), pp. 353–358.