Re-Vibe: Vibration-based Indoor Person Re-Identification through Cross-Structure Optimal Transport

Yiwen Dong*
ywdong@stanford.edu
Stanford University
Stanford, California, USA

Jiacheng Zhu
Carnegie Mellon University
Pittsburgh, USA

Hae Young Noh
Stanford University
Stanford, USA

ABSTRACT

Indoor person re-identification (i.e., matching the same person across various locations) is important in making smart building applications more ubiquitous for future workspaces. Specific examples include occupant tracking, energy flow optimization, and security surveillance. Typical person re-identification (re-ID) systems rely on cameras to match the same person across different locations. While it is accurate, vision-based sensing requires direct line-of-sight and may raise privacy concerns. Besides, other sensing methods (e.g., wearables, pressure mats, RF/WiFi) are restricted to requirements such as carrying devices, dense deployment, and motion through wave path. In this paper, we develop *Re-Vibe*, the first system that re-identifies people through footstep-induced floor vibrations. Our approach is device-free, wide-ranged, and perceived as more privacy-friendly. The main challenge in developing this system is the discrepancy in structural properties and variations in footsteps at distinct locations, leading to different vibration feature patterns within the same person. As a result, it is difficult to match the same person's footsteps across various locations. To overcome this challenge, we model the structure and human influence on the vibration signals to develop a cross-structure transfer model that enables footstep matching despite the structure and human variations. The model is formulated based on optimal transport (OT), which compares the distribution pattern of a person's footstep features across multiple structures for person re-ID. We evaluate *Re-Vibe* at two locations in a building with wooden and concrete flooring. *Re-Vibe* achieved a 0.9 F-1 score with four people in person re-ID across these two locations.

CCS CONCEPTS

• **Human-centered computing → Ubiquitous and mobile computing**.

KEYWORDS

person re-identification, human-building interaction, optimal transport structural vibration, unsupervised domain adaptation

ACM Reference Format:
Yiwen Dong, Jiacheng Zhu, and Hae Young Noh. 2022. Re-Vibe: Vibration-based Indoor Person Re-Identification through Cross-Structure Optimal Transport. In *The 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '22), November 9–10, 2022, Boston, MA, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3563357.3566134

1 INTRODUCTION

Re-identifying (re-ID) a person across various locations enables more ubiquitous occupant sensing as people move around in the building, which is important for many smart building application scenarios such as the future of work [10]. Specific examples include occupant tracking for environmental hazard proximity tracking [2, 14], energy flow optimization for human comfort, security surveillance for access control [1, 19]. Compared to person identification, re-ID does not require pre-collected labels at each sensing location [23], which improves the scalability and flexibility of the smart sensing devices to be deployed over multiple locations.

The existing person re-ID system uses non-overlapping cameras to match the same person across different locations, which has achieved high accuracy in various benchmark datasets [12]. However, the camera requires direct line-of-sight, and installing them in private building spaces often raises privacy concerns. Other approaches (including wearables, pressure-, RF-based sensing) overcome these limitations [7, 9], but they are often restricted in operational requirements, such as carrying devices, dense deployment, and constrained walking paths. Vibration-based sensing, however, only requires sparsely deployed sensors, is obstruction-insensitive, wide-ranged, and perceived as more privacy-friendly [17]. The primary insight of this approach is that people's footstep forces generate floor vibrations, which are consistent within the same person across different locations. In prior studies, footstep-induced floor vibrations have been shown to be effective in person identification and new person discovery [4, 5, 18]. However, it requires training data for each new location and is unable to match the same person across multiple locations. While prior work has succeeded in cross-structure transfer using transfer component analysis [15], it is limited to footstep detection and is unable to distinguish between people. This is because the method focuses on the difference between various structures while neglecting the difference in people's footstep patterns within each structure.

In this paper, we introduce *Re-Vibe*, a system that aims to re-identify people across various locations in a building. *Re-Vibe* is, to the best of our knowledge, the first vibration-based person re-ID system. *Re-Vibe* first extract features from the vibration signals as a person pass by the sensors at different locations. Then, it learns

---

BuildSys '22, November 9–10, 2022, Boston, MA, USA

ACM ISBN 978-1-4503-9890-9/22/11…$15.00
https://doi.org/10.1145/3563357.3566134

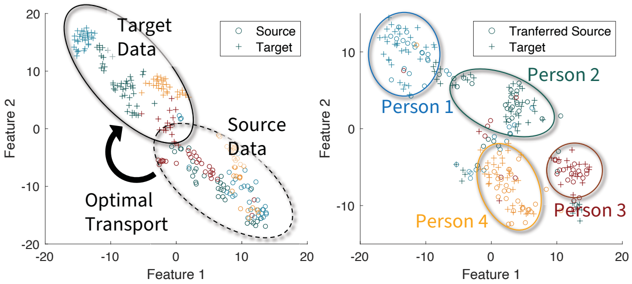

Figure 1: 2D T-SNE plot for footsteps before (left) and after (right) the transfer. Different color indicates different people. OT effectively aligns each person's footstep distribution based on their geometric relationships in the feature space.

a function that transfers the footstep features from the source location (i.e., the location where the person first appears) to the target location (i.e., a new location where the person appears later) by considering the distribution differences among people's footstep features at each given location. This function minimizes the overall discrepancy between the footstep data of the source and target locations while distinguishing different people's footstep features at each location. Finally, features from the source location are transferred to the target location. Given that the labels are first assigned to each person at the source location, the same person is re-identified at the target location based on the transferred source location data (as described in Figure 1). Compared to the existing cross-structure transfer method [15], our approach has the advantage of modeling each person's footstep distributions on top of reducing the difference across various locations.

It is challenging to re-ID people across various locations, mainly due to the difference in structural properties and footsteps over these locations. First, the structural properties at different locations vary, resulting in distinct patterns in footstep-induced vibration signals for each location (see Figure 2). It is difficult to match the same person across dissimilar signal patterns. Secondly, the same person's footstep forces and the number of recorded footsteps also vary across locations. Specifically, the same person's footsteps' characteristics vary due to psychological and physical variations when walking around in the building. Moreover, there are footstep "outliers" due to tripping or dragging when walking, making it challenging to extract representations of individuals' walking patterns. In addition, the number of samples from the source and the target location is often imbalanced. A location may only have limited samples to represent a person's footstep pattern, making it difficult to match with another location that has much more samples.

To address these challenges, we model the structure and human influences on the vibration signals to develop a cross-structure transfer that contains both structure and human information. Our approach is formulated based on the Optimal Transport (OT) theory [21] because OT considers both the distribution of each person [13, 24] and the distance between structures, which aligns well with our re-ID purpose. Therefore, variations in footstep characteristics are taken into account, and the number of samples at each location is flexible.

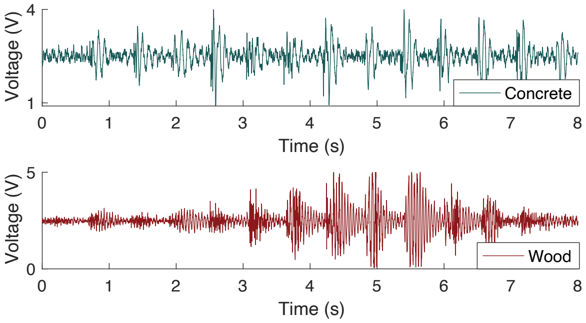

Figure 2: Footstep-induced floor vibrations of the same person walking at two locations, with concrete and wooden floor, respectively.

We evaluate *Re-Vibe* through a field experiment with four people walking at two locations with wooden and concrete floors, respectively. Our system achieved a 0.9 F-1 score in person re-ID.

The main contributions of this work are:

- We introduce the first system to re-identify a person across different structures using footstep-induced floor vibrations.
- We model the structure and human influence on the vibration signals based on the Optimal Transport (OT) theory in the context of linear and non-linear dynamical systems.
- We conduct real-world evaluation across concrete and wooden structures, which achieves a 0.9 F-1 score and 2× error reduction compared with the existing methods.

In the rest of the paper, we first model the structural and footstep variations in vibration signals based on the OT theory (Section 2), then provide an overview of the system (Section 3). A field experiment and the results are then discussed (Section 4). The remaining challenges are summarized and a conclusion is drawn at the end (Section 5, Section 6).

2 MODELING THE FOOTSTEP-INDUCED VIBRATIONS ACROSS TWO STRUCTURES

We analytically model the relationship between the footstep-induced vibrations from the same person across two structures, based on the OT theory and the representations of dynamical systems. The dynamical system consists of input (i.e., the footstep forces) and output (i.e., the vibration signals). The relationship between the input and the output is determined by the characteristics of the structure. To this end, we formulate this relationship based on the OT theory with the dynamical system under both linear and non-linear assumptions.

Optimal Transport Theory: To begin with, the objective of optimal transport is to find an optimal coupling matrix with elements π_{ij} that matches sample i in the source domain to the sample j in the target domain through a transport map function $T(\cdot)$ [25], in order to minimize the Wasserstein distance, which represents the similarity between the target domain data and the transferred source domain data [21, 26]:

$$T(\cdot), \pi_{ij} = \arg\min_{\pi,T} \sum_{i,j} d(T(X_i^s), X_j^t)\pi_{ij} \tag{1}$$

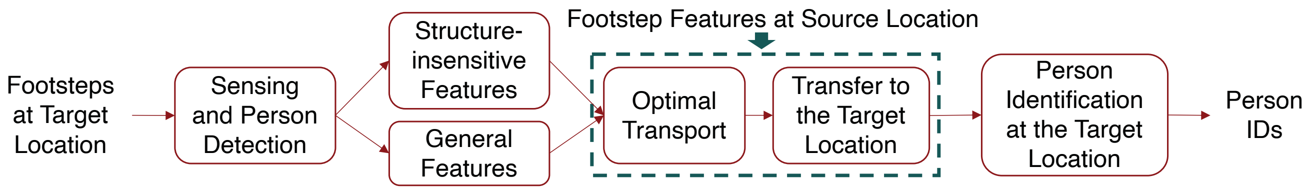

Figure 3: *Re-Vibe* System Overview

where X_i^s, X_j^s denotes footstep force vectors at source and target domains to represent the distribution of a person's footsteps. Therefore, $d(T(X_i^s), X_j^t)$ represents a distance measure between the transferred source and the target.

Linear Dynamical Systems: In a linear time-invariant (LTI) system H_k, the relationship of footstep forces X_i and the structural responses Y_i in the frequency domain can be represented as [15]:

$$Y_i = H_k X_i \tag{2}$$

Consider a special case when $d(\cdot, \cdot)$ is initialized as euclidean distance and the transfer function is a linear kernel $T(Y_i) = W^T Y_i$, the objective function can be written as:

$$min \sum_{i,j} \left\| W^T H_s X_i^s - H_t X_j^t \right\|_2 \pi_{ij} \tag{3}$$

As the same person's gait patterns are relatively consistent when walking at different locations [18], we further simplify the expression by assuming that a person's footstep forces share the same distribution across two structures (i.e., $X_i^s \overset{\mathrm{d}}{=} X_j^t, \forall i, j$). Then an obvious solution to the above objective is:

$$H_t = W^T H_s \tag{4}$$

In this case, the linear relationship between the source and target structures can be sufficiently represented by the parameters W in the transfer function. Moreover, the physical meaning behind the linear combination of frequency domain features can be regarded as the scaling and stretching of floor vibration signals in the time domain, according to the stretch theorem in the Fourier transform and the property of linear systems.

Non-linear Dynamical Systems: In a non-linear time-invariant (NLTI) system $H_k(\cdot)$, the relationship of footstep force distribution X_i and the structural response Y_i in the frequency domain is written by:

$$Y_i = H_k(X_i) \tag{5}$$

Then, the objective function can be generalized as follows:

$$\min_{H_s, H_t} \sum_{i,j} d(T(H_s(X_i^s)), H_t(X_j^t))\pi_{ij} \tag{6}$$

Following the same assumption that a person has the same footstep force distribution exerting to the floor (denoted as X_p in general), an obvious solution to the objective function is:

$$T(H_s(X_p)) = H_t(X_p) \tag{7}$$

In this case, the transport map function $T(\cdot) = H_t(H_s^{-1}(\cdot))$ represents a general relationship between the source and target dynamical systems, which can be interpreted by a composition of an inverse source system followed by a forward target system.

3 VIBRATION-BASED PERSON RE-ID THROUGH OPTIMAL TRANSPORT

As shown in Figure 3, *Re-Vibe* consists of four steps. When people walk by the sensors at the target structure, floor vibrations are captured to determine the presence of a person. Then, structure-insensitive and general features are extracted to represent the people's walking patterns. In the main step (boxed with green dotted line), we learn the optimal transport map based on the recorded footsteps from the source and target locations. After that, footstep features are then transferred to the target location. Finally, we conduct person identification at the target location to determine the owners of the collected footsteps.

3.1 Sensing and Person Detection

In order to capture footstep-induced floor vibrations, an array of geo-phones are mounted at the side of the corridor and passively receive vibration signals. When the sensor readings exceed 3× the threshold of white noise, the system is activated. The footstep impulses are first differentiated from the other source of vibrations (e.g., ball dropping, door closing). More than four consecutive impulsive events that have less than three standard deviations of duration and interval between them are detected as footsteps, which represent the presence of a person.

3.2 Structure-Insensitive Feature Extraction

We extract general footstep features (i.e., mean signal amplitudes in every 10 Hz from 0-200 Hz) and structure-insensitive features (such as step time, left and right footstep symmetry) to represent a person's gait [6, 8]. Since the latter is computed based on the time difference between footsteps, they depend less on the structural properties and thus keep the same distribution of a person's footstep data across different locations.

3.3 Cross-Structure Optimal Transport

With features extracted from footsteps, we use optimal transport described in Section 2 to learn the transport map function between source and target locations. To overcome the imbalanced footstep problem, each target location footstep trace is matched with several source domain footstep traces (or vice versa) through the Sinkhorn algorithm [3]. Each pair is assigned different weights during training, depending on how similar they are in the data distribution of their domain. The more similar two footstep traces are, the higher weights they are likely to be assigned. This ensures that footsteps more representative of their owners are weighted more than the less representative ones (i.e., outliers) when estimating the transport map function. The transport map is parameterized as the Gaussian

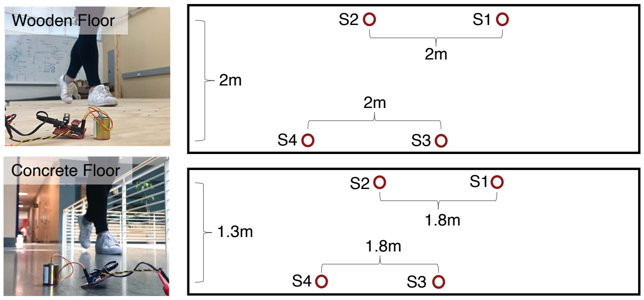

Figure 4: *Re-Vibe* Experiment Setup: each person walks at two different locations, one with a wooden floor and the other with a concrete floor.

kernel, which is smooth, fast in convergence, and flexible to represent infinitely complex models [11]. Parameters that describe the transfer are then learned by minimizing the gap between the source and target location data through an iterative optimization [20].

After obtaining the transport map function, footstep data recorded in the source location are then transformed to the target location, the labels of which are used for person re-identification.

3.4 Person Re-ID at the Target Location

To re-identify people at the target location, a person identification algorithm developed by our prior work is used [18]. The training set contains transferred source domain data and their labels. Correspondingly, the newly collected footsteps at the target location serve as the test set. The algorithm is based on support vector machines with radial basis function kernels (SVM-RBF) for its interpretability through confidence scores and better performance over the baseline models. The classification result is the identity of the person to whom the footstep traces belong.

4 FIELD EVALUATION

To evaluate *Re-Vibe*, we design a walking experiment with four participants across two locations in a building. (approved Stanford IRB No. 54912)

4.1 Experiment Setup

The experiment was conducted at two locations with wooden and concrete floors. As shown in Figure 4, four SM-24 geophones were mounted on the surface of the floor [22]. As the concrete floor has higher stiffness and less deformation than the wooden floor structure, the velocity was amplified by 1000× and 200×, respectively, to improve the signal-to-noise ratio. Each participant walked across two locations with their normal gaits. Their footstep-induced floor vibrations are captured and converted by National Instrument DAQ from analog to digital signals. The sampling rate is 26.5 kHz. Due to the difference in damping of the structures, the sensing range is around 10 meters and 3 meters for wooden and concrete floors, respectively.

4.2 System Performance

The system has achieved a 0.9 F-1 score in person re-identification during the preliminary testing with four people, which has a 2× error reduction compared with the baseline (assumes the same weight for all pairs) [15]. We also observe that the accuracy drops when more people join the system because footstep features are less separable with an increasing number of occupants.

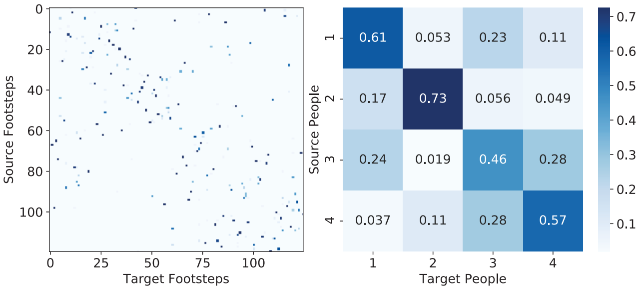

Figure 5: Weights of source-target pairs across samples (left) and people (right). The same person's footstep samples have higher weights (darker in color) because they share similar characteristics.

Effect of Structure-Insensitive Features: We observed that time-domain structure-insensitive features contribute significantly to the overall prediction in addition to the frequency-domain general footstep features developed in previous studies [5]. We investigate the importance of features through a random forest classifier and found that these features represent 26% of importance among all features.

Interpreting the Coupling Matrix: As mentioned in Section 3, the coupling matrix shows how the weights are assigned between source-target pairs. As we can observe in Figure 5, more weights are assigned to footstep pairs within the same person, while the coupling between different people is weaker unless their footsteps share similar features with each other.

5 FUTURE WORK

With the promising results in the evaluation and the rigorous theoretical foundation of optimal transport, there are many directions to extend the work in the future.

Out-of-Sample Footstep Data Modeling: In real-life applications, out-of-sample footstep data (e.g., a newcomer's footsteps, distribution shift of an existing people's footsteps) may appear [16]. Our prior studies have developed a system based on Dirichlet Process (DP) to model previously unobserved people [5], and our ongoing works are exploring how walking speed, different types of shoes, and emotional changes affect the footstep-induced structural vibrations. We will further explore how *Re-Vibe* can be extended to deal with walking pattern variations under scenarios such as emergency evacuation, stress level changes, and potential data shifts caused by shoe types and gait disorders.

Footstep Data Augmentation: It is labor-intensive and cost-inefficient to collect and label footsteps at every sensing location. Therefore, it would be more efficient if we could augment a small amount of data and labels obtained at one source location to multiple target locations over the entire building. The augmentation can be achieved by learning the physical relationship between the various locations from the OT transformation [26].

Footstep Association from Mis-matched Subjects: In real-life scenarios, it is not realistic to record everyone's data at all locations. In that case, we can improve the Sinkhorn algorithm in OT such that the system is robust to the absence of persons at any target location. The ultimate goal is to match the same person's footsteps even when no label is available at any location (i.e., zero-shot domain adaptation).

6 CONCLUSION

In conclusion, we introduced *Re-Vibe*, the first vibration-based person re-ID system through cross-structure optimal transport. *Re-Vibe* leverages the physical insights of the footstep-induced structural responses based on the formulation of optimal transport. Our approach overcomes the challenge of the discrepancy in structural properties and variations in footsteps at distinct locations by analytically modeling the structure and human influence on the vibration signals. We also take a probabilistic perspective to accommodate the imbalanced sample number and footstep outliers across various locations. We conduct a real-world walking experiment on two different types of floors. *Re-Vibe* achieved a 0.9 F-1 score with four people and a 2× error reduction compared with previous work.

ACKNOWLEDGMENTS

This research was supported in part by the National Science Foundation (under grant NSF-CMMI-2026699). Special thanks to Megan Zhang, Jingxiao Liu, Helen Zhang, Qian Dong, Simba Wu, Yueming Zhuo, Rachel Yan, Jerry, Maggie and Thomas for their kind support with the field experiment.

REFERENCES

[1] Le An, Xiaojing Chen, Mehran Kafai, Songfan Yang, and Bir Bhanu. 2013. Improving person re-identification by soft biometrics based reranking. In *2013 seventh international conference on distributed smart cameras (ICDSC)*. IEEE, 1–6.

[2] Long Chen, Haizhou Ai, Zijie Zhuang, and Chong Shang. 2018. Real-time multiple people tracking with deeply learned candidate selection and person re-identification. In *2018 IEEE international conference on multimedia and expo (ICME)*. IEEE, 1–6.

[3] Marco Cuturi. 2013. Sinkhorn Distances: Lightspeed Computation of Optimal Transport. In *Advances in Neural Information Processing Systems*, C.J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger (Eds.), Vol. 26. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2013/file/af21d0c97db2e27e13572cbf59eb343d-Paper.pdf

[4] Yiwen Dong, Jonathon Fagert, Pei Zhang, and Hae Young Noh. 2021. Non-parametric Bayesian Learning for Newcomer Detection using Footstep-Induced Floor Vibration. In *Proceedings of the 20th International Conference on Information Processing in Sensor Networks (co-located with CPS-IoT Week 2021)*. 404–405.

[5] Yiwen Dong, Jonathon Fagert, Pei Zhang, and Hae Young Noh. 2023. Stranger Detection and Occupant Identification Using Structural Vibrations. In *European Workshop on Structural Health Monitoring*. Springer, 905–914.

[6] Yiwen Dong, Joanna Jiaqi Zou, Jingxiao Liu, Jonathon Fagert, Mostafa Mirshekari, Linda Lowes, Megan Iammarino, Pei Zhang, and Hae Young Noh. 2020. MD-Vibe: physics-informed analysis of patient-induced structural vibration data for monitoring gait health in individuals with muscular dystrophy. In *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*. 525–531.

[7] Kayne Duncanson, Simon Thwaites, David Booth, Ehsan Abbasnejad, William SP Robertson, and Dominic Thewlis. 2021. The Most Discriminant Components of Force Platform Data for Gait Based Person Re-identification. (2021).

[8] Jonathon Fagert, Mostafa Mirshekari, Shijia Pan, Pei Zhang, and Hae Young Noh. 2017. Characterizing left-right gait balance using footstep-induced structural vibrations. In *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2017*, Vol. 10168. SPIE, 357–365.

[9] Lijie Fan, Tianhong Li, Rongyao Fang, Rumen Hristov, Yuan Yuan, and Dina Katabi. 2020. Learning longterm representations for person re-identification using radio signals. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10699–10709.

[10] Marylène Gagné, Sharon K Parker, Mark A Griffin, Patrick D Dunlop, Caroline Knight, Florian E Klonek, and Xavier Parent-Rocheleau. 2022. Understanding and shaping the future of work with self-determination theory. *Nature Reviews Psychology* (2022), 1–15.

[11] Ziv Goldfeld and Kristjan Greenewald. 2020. Gaussian-smoothed optimal transport: Metric structure and statistical efficiency. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 3327–3337.

[12] Mengran Gou, Ziyan Wu, Angels Rates-Borras, Octavia Camps, Richard J Radke, et al. 2018. A systematic evaluation and benchmark for person re-identification: Features, metrics, and datasets. *IEEE transactions on pattern analysis and machine intelligence* 41, 3 (2018), 523–536.

[13] William Han, Jielin Qiu, Jiacheng Zhu, Mengdi Xu, Douglas Weber, Bo Li, and Ding Zhao. 2022. An Empirical Exploration of Cross-domain Alignment between Language and Electroencephalogram. *arXiv preprint arXiv:2208.06348* (2022).

[14] John Howard, Vladimir Murashov, Emanuele Cauda, and John Snawder. 2022. Advanced sensor technologies and the future of work. *American Journal of Industrial Medicine* 65, 1 (2022), 3–11.

[15] Mostafa Mirshekari, Jonathon Fagert, Shijia Pan, Pei Zhang, and Hae Young Noh. 2020. Step-level occupant detection across different structures through footstep-induced floor vibration using model transfer. *Journal of Engineering Mechanics* 146, 3 (2020).

[16] Debarghya Mukherjee, Aritra Guha, Justin M Solomon, Yuekai Sun, and Mikhail Yurochkin. 2021. Outlier-robust optimal transport. In *International Conference on Machine Learning*. PMLR, 7850–7860.

[17] Shijia Pan, Ningning Wang, Yuqiu Qian, Irem Velibeyoglu, Hae Young Noh, and Pei Zhang. 2015. Indoor person identification through footstep induced structural vibration. In *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications*. 81–86.

[18] Shijia Pan, Tong Yu, Mostafa Mirshekari, Jonathon Fagert, Amelie Bonde, Ole J Mengshoel, Hae Young Noh, and Pei Zhang. 2017. Footprintid: Indoor pedestrian identification through ambient structural vibration sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–31.

[19] Shanchen Pang, Sibo Qiao, Tao Song, Jianli Zhao, and Pan Zheng. 2019. An improved convolutional network architecture based on residual modeling for person re-identification in edge computing. *IEEE Access* 7 (2019), 106748–106759.

[20] Michaël Perrot, Nicolas Courty, Rémi Flamary, and Amaury Habrard. 2016. Mapping estimation for discrete optimal transport. *Advances in Neural Information Processing Systems* 29 (2016).

[21] Jian Shen, Yanru Qu, Weinan Zhang, and Yong Yu. 2018. Wasserstein distance guided representation learning for domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.

[22] Sparkfun. [n.d.]. *SM-24 Geophone Element*. Input/Output, Inc.

[23] Liang Zheng, Yi Yang, and Alexander G Hauptmann. 2016. Person re-identification: Past, present and future. *arXiv preprint arXiv:1610.02984* (2016).

[24] Jiacheng Zhu, Gregory Darnell, Agni Kumar, Ding Zhao, Bo Li, Xuanlong Nguyen, and Shirley You Ren. 2022. PhysioMTL: Personalizing Physiological Patterns using Optimal Transport Multi-Task Regression. In *Conference on Health, Inference, and Learning*. PMLR, 354–374.

[25] Jiacheng Zhu, Aritra Guha, Dat Do, Mengdi Xu, XuanLong Nguyen, and Ding Zhao. 2021. Functional optimal transport: map estimation and domain adaptation for functional data. *arXiv preprint arXiv:2102.03895* (2021).

[26] Jiacheng Zhu, Jielin Qiu, Zhuolin Yang, Douglas Weber, Michael A Rosenberg, Emerson Liu, Bo Li, and Ding Zhao. 2022. GeoECG: Data Augmentation via Wasserstein Geodesic Perturbation for Robust Electrocardiogram Prediction. *arXiv preprint arXiv:2208.01220* (2022).