Augmenting Vibration-Based Customer-Product Interaction Recognition with Sparse Load Sensing

Yue Zhang
yzhang58@ucmerced.edu
University of California, Merced
Merced, CA, USA

Shiwei Fang
shiweifang@cs.umass.edu
University of Massachusetts Amherst
Amherst, MA, USA

Carlos Ruiz
carlos@aifi.com
AiFi Inc.
Burlingame, CA, USA

Zhizhang Hu
zhu42@ucmerced.edu
University of California, Merced
Merced, California, USA

Shubham Rohal
srohal@ucmerced.edu
University of California, Merced
Merced, CA, USA

Shijia Pan
span24@ucmerced.edu
University of California, Merced
Merced, CA, USA

ABSTRACT

This paper introduces a multimodal solution for autonomous retail customer-product interaction recognition using a combination of vibration and load sensing. Scalable and robust customer-product interaction recognition is important for autonomous retail. Current efforts focus on computer vision-based approaches, which are prone to occlusion from both customers and shelves. A densely deployed load cell array can mitigate this issue, however, the high cost and maintenance effort make it difficult to scale up for large deployments. Vibration-based approaches are also explored to detect such interaction, however, the robustness over noisy environments is limited.

We propose a multimodal solution with sparse vibration and load sensing on the shelves. These two modalities are complementary in terms of information – load sensing can effectively detect the weight changes on the shelf while vibration sensing can recognize detailed interaction. Preliminary results are presented to demonstrate the complementarity of these two modalities. Our system is able to augment product recognition performance with the combination of sparse vibration and load sensing.

CCS CONCEPTS

• **Computer systems organization → Sensors and actuators**.

KEYWORDS

Vibration sensing; load sensing; multimodal system

ACM Reference Format:
Yue Zhang, Shiwei Fang, Carlos Ruiz, Zhizhang Hu, Shubham Rohal, and Shijia Pan. 2023. Augmenting Vibration-Based Customer-Product Interaction Recognition with Sparse Load Sensing. In *Cyber-Physical Systems and Internet of Things Week 2023 (CPS-IoT Week Workshops '23), May 09–12, 2023, San Antonio, TX, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3576914.3589560

CPS-IoT Week Workshops '23, May 09–12, 2023, San Antonio, TX, USA

ACM ISBN 979-8-4007-0049-1/23/05…$15.00
https://doi.org/10.1145/3576914.3589560

1 INTRODUCTION

Autonomous retail utilizes sensors to acquire information on the interaction between customers and products and achieves intelligent inventory monitoring, as well as cashier-less checkout [3, 7, 18]. Current prevailing solutions are mostly focusing on computer vision [4, 8], which is sensitive to the occlusion that comes from both shelves and customers in the environment. Therefore, many sensing modalities are investigated as complementary solutions to computer vision or standalone solutions. For example, RFID tags are placed on each product for identification and tracking [16, 17]. It is low cost to install for individual products, however, the installation for all the products in large-scale retail stores becomes expensive, in terms of both labor and cost. Load sensor arrays are then developed, as a complementary modality to computer vision [18], which can effectively detect products by their weights without the requirement of the Line-of-Sight (LoS). However, it requires dense deployment (e.g., 24 sensors per shelf), which poses challenges for cost, installation, and maintenance. The vibration-based approach has been explored to reduce the need for dense deployments [20], as the acoustic signals propagate in solid and contain rich information. However, this approach faces challenges from ambient acoustic sources in the environment.

To effectively and robustly detect and recognize the product that the customer interact with, we combine load and vibration sensing, as they are complementary in detection and recognition performance. The intuition is twofold. First, load sensing is sensitive to the weight changes on the shelf only, therefore its detection accuracy is high [18]. However, an accurate detection requires a dense deployment, which is costly. We design a shelf loading sensing scheme that can effectively detect the weight change with low-cost and sparse deployed pressure sensors. Second, the customer-product interaction (pick-up or put down a product) causes the shelf to vibrate, and this vibration signal contains rich information. For example, products with different textures (glass v.s. plastic) may excite the surface on different frequency bands, resulting in different vibration waveforms [6]. However, there are many ambient vibration sources that can be detected by the vibration sensor, which pose challenges for accurate detection. Therefore, these two modalities are effective on different tasks.

We combine these two modalities with a task-level fusion, where the low-cost load sensing is used for detecting the events, and

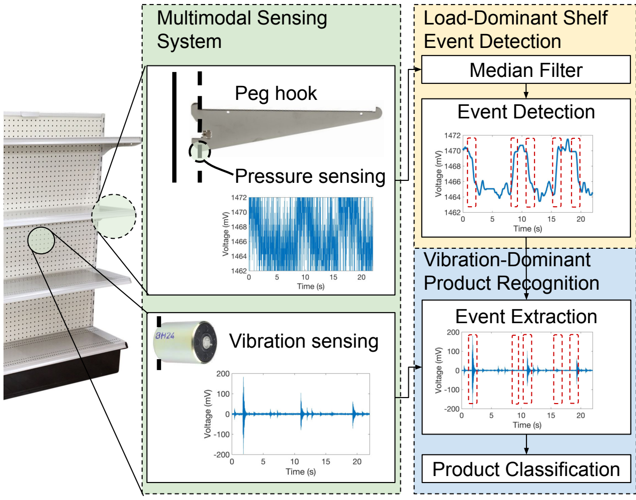

Figure 1: Illustration of our multimodal sensing system that combines vibration sensing with low-cost load sensing.

the vibration sensing is used for recognizing the products customers interact with. To verify the proposed multimodal scheme, we implement the system and conduct real-world experiments. Our preliminary analysis demonstrates a 2.6× and 1.6× improvement for the product interaction detection and recognition performance respectively, compared to the single modality-based approach. Our contributions are summarized as follows.

- We introduce a low-cost vibration and load sensing combination for robust customer-product interaction recognition at autonomous retails.
- We design a shelf load sensing scheme leveraging the existing retail shelf's structure (the right angle lever structure in a peg hook) and two low-cost pressure sensors.
- We identify the complementary performance of the noisy vibration and load sensing, and propose a task-level multimodal fusion scheme that mitigates noises' impacts in both modalities.
- We conduct real-world experiments and preliminary analysis on our design.

2 RELATED WORK

This section lists sensing modalities explored for autonomous retail and compares their advantages and disadvantages.

Computer Vision Only. Computer vision is the most widely used solution in autonomous retail nowadays [4, 8]. In most cases, multiple overhead cameras are used to track customer and their activities in the store [3, 7]. With the development of posture tracking [5] and image recognition [15], they often can achieve high accuracy when the Line-of-Sight (LoS) is guaranteed. However, in real-world scenarios, it often requires multiple cameras to cover one area from different angles to ensure that, which results in a high cost of computation and maintenance. Apart from this, this performance is also impacted by the lighting condition in the store and the distance of interaction from the camera.

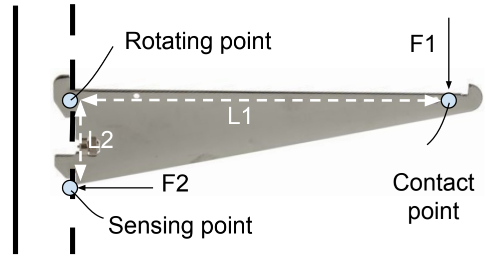

Figure 2: Converting shelf into low-cost load sensor with sparsely placed pressure sensors.

RFID Only. RFID, on the other hand, has been explored to detect events or customer interaction with products in the store [16, 17]. Individual RFID tags are placed on each product and are scanned by the scanner placed on the shopping carts, shelves, or checkout counter to identify the products. Since the tag provides a one-to-one mapping to the product identity, the recognition accuracy is high. However, despite the low cost of each tag, the total cost of the tag and labeling quickly accumulates when the number of products scales up.

Shelf Load Sensing. To compensate for the disadvantages of vision and RFID in autonomous retail settings, load-based sensing has been explored. Multiple load cells are placed underneath the shelves to capture the weight changes on the shelve caused by product pick-up and putting down [9, 10, 13]. This approach produces accurate product recognition and interaction detection results when only one product is placed on top by identifying the weight difference. However, it fails to recognize item(s) when the same amount of weight changes are caused by different product interactions.

Multimodal Fusion. Many multimodal approaches have been explored to the overcome the challenges presented by each modality. For example, vision plus load/weight sensing have been explored to overcome the occlusion problem presented by vision, while still leveraging its effective customer tracking [11, 18]. Apart from this, the vision plus RFID approach has also been explored [12]. Although they overcome the vision's occlusion problem, they require dense deployment and the retrofitting cost is still high. In this work, we explore fine-grained multimodal sensing on the shelf, which provides a low-cost solution that can be further fused with existing vision-based approaches.

3 SYSTEM DESIGN

The system consists of three major modules – multimodal sensing, load-dominant event detection, and vibration-dominant product recognition – as shown in Figure 1. First of all, we design a low-cost and scalable load sensing scheme based on the lever structure in existing retail shelves. Second, we leverage the complementarity of load sensing and vibration sensing to design a cross-modal event detection method that is robust to ambient noises. Finally, we leverage the informative vibration signal segments for accurate product recognition.

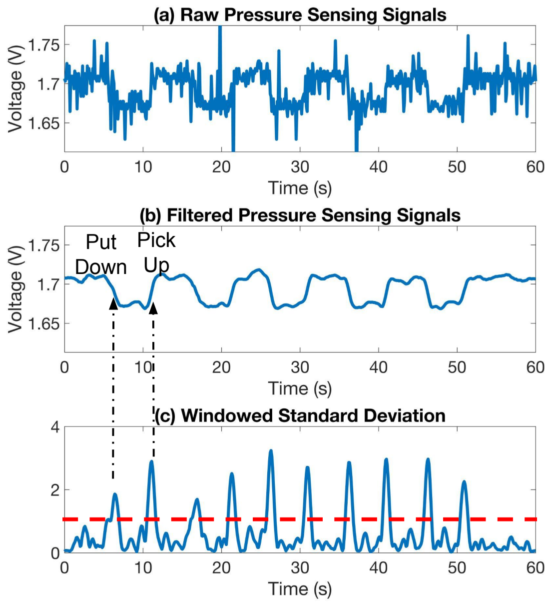

Figure 3: Pressure sensor outputs and event detection. (a) depicts raw pressure sensor signals when a soda can is picked up and put down on the shelf multiple times. (b) shows the filtered signals after the median filter and the moving average filter. (c) shows the windowed standard deviation values calculated from the filtered signal.

3.1 Converting Shelf for Low-Cost Load Sensing

To effectively detect product interaction, we convert the retail shelf into a load sensor leveraging its structural characteristics. The retail shelf, a.k.a. gondola, often uses the peg hook design for the shelf installation on the frame/back panel, as shown in Figure 2. The peg hook structure acts as a right-angle lever when force is applied to the shelf, depicted as L1. The output arm of the right angle is against the frame or the back panel of the shelf. The force applied to the lever can be modeled with equation $F_1L_1 = F_2L_2$. Therefore, when a product is placed on the shelf with a contacting point L_1 distance away from the rotating point, the force F_1 applied will cause the force $F_2 = F_1L_1/L_2$ applied at the sensing point as shown in Figure 2. Since the structure of the peg hook has a structure with $L_1 >> L_2$, sensing at the output arm of the right-angle lever would results in a higher signal-to-noise ratio. Therefore, this allows us to use a low-cost (noisy) pressure sensor to capture the load change.

3.2 Load-Dominant On-Shelf Event Detection

The load change on the shelf directly reflects the interaction between the customer and the product – whether it is a pick-up that reduces the pressure at the sensing point or a put-down that increases the pressure at the sensing point. Figure 3 (a) depicts the raw pressure sensor's signal when a product (soda can, 380 grams) is picked up and put down multiple times in one minute. We can observe that the trend of the changes in the signals matches with the load change (pick up and put down) on the shelf.

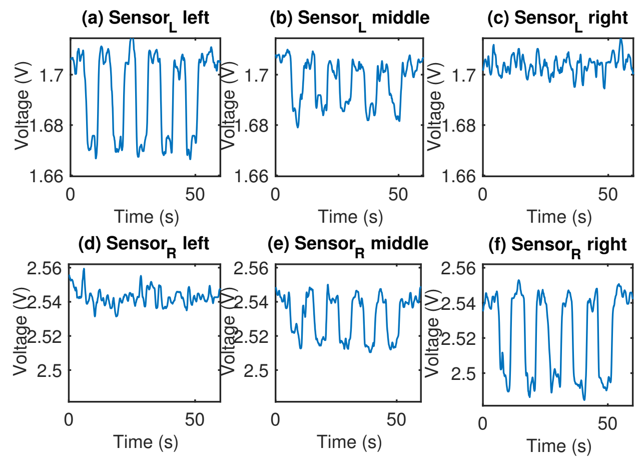

Figure 4: Filtered signals from pressure sensors on two sides of the shelf. (a-c) shows signals from the pressure sensor installed on the left side of the shelf, when the customer interacts with the product at left side, middle, and right side of the shelf. (d-e) shows the corresponding signals from the sensor on the right side.

Filtering. Because the raw signal contains a strong salt and pepper noise, and white noise. We use a median filter and a moving average filter to denoise the signal. Figure 3 (b) shows the filtered signal, where the pressure changes and stabilization trends are observable.

Single-Sensor Detection. To efficiently extract interaction events that cause shelf load changes, we first apply a sliding window of size W on the filtered signal and calculate the standard deviation of the values in this window as σ [18]. This sliding window will output an array of values that reflect the pressure variation in time, as shown in Figure 3 (c). For the k^{th} window, if σ_k is larger than a threshold Th_σ, we consider that an event occurs. For consecutive windows with $\sigma_k > Th_\sigma$, we consider them as part of the same event and select the timestamp of the k^th window with the highest σ_k value as the i^{th} event's time T_i. The pressure sensor values at the start and the end of the event indicate the event category – whether it is a pick-up or put-down of the product – as an example shown in Figure 3 (b).

Shelf-Level Detection. Since the pressure sensor measures the load on the side of the shelf leveraging the peg hook structure, this sensitivity across the shelf varies. Figure 4 shows the filtered signals from sensors on two sides of the shelf. (a-c) depict signals captured by the sensor on the left side of the shelf, and (d-e) show signals from the sensor on the right side, with the customer interacting on the left side, middle, and right side of the shelf. We observe that events occurring at locations close to the sensor, e.g., signals in (a), show a higher signal-to-noise ratio (SNR) in terms of the voltage changes caused by load changes. Therefore, for each shelf, we install pressure sensors on both sides of the shelf-frame contacting point to efficiently capture load changes at different locations on the shelf. We merge these two sensors detected events for robust detection. For these two pressure sensors, they independently detect N and M events at $T_i^1, i = 1...N$ and $T_j^2, j = 1...M$. Assuming only one event occurs on the shelf at a time, we consider when $|T_i^1 - T_j^2| < Th_{Event}$ as the same event, and report the event time as the one with a higher σ value.

3.3 Vibration-Dominant Product Recognition

The vibration sensor is placed on the back panel of the retail shelves to efficiently acquire the vibration signals induced by the customer-product interaction [20]. However, as the acoustic signals often contain information from multiple ambient sources, e.g., people passing by or a cart running into the shelf, it is difficult to model all the vibration sources in the environment. Therefore, our system uses the event detected by the load sensing to extract the corresponding vibration signal segments for recognition. First, we conduct an energy-based event detection [20] on the vibration data, and output the timestamp T_j^{vib} of all the detected potential events. We consider the vibration sensing and load sensing detect the same event when $|T_i - T_j^{vib}| < Th_{align}$, and outputs vibration signal segments for all qualified T_j^{vib}.

For each extracted signal segment, we first normalize its signal energy, then extract the frequency components' amplitude as the feature of this event signal. A classifier is trained with vibration signal segment and product identity labels. In this study, we adopt the support vector machine (SVM) model [19] with a radial basis function (RBF) kernel as the classifier for its strong empirical performance in vibration signal-based classification tasks [20]. To tune the hyperparameters of the SVM model, we conduct k-fold cross-validation grid search. The searching space consists of the SVM model's regularization parameter (referred to as "C") and the RBF kernel coefficient (referred to as "γ"). The selected hyperparameters are the combination that maximizes the performance metric in the left-out data after the k-fold split [14].

4 IMPLEMENTATION AND PRELIMINARY RESULTS

We implement the system and acquire real-world data with three products of small sizes to demonstrate the feasibility of our multi-model scheme.

4.1 Experiment Setup

Figure 5 depicts the experimental setup. The pair of pressure sensors are installed on two sides of the shelf marked as $Sensor_L$ and $Sensor_R$ in Figure 5 (a) and (b). The vibration sensor, marked as $Sensor_{vib}$, is mounted on the center of the back panel. Figure 5 (d) shows the vibration sensor geophone [2] installed on the back panel of the retail shelves. The load sensing setup is shown in Figure 5 (e), where a pressure sensor [1] is placed between the frame and the peg hook's right angle lever output arm. We test the multimodal scheme for interaction detection and recognition over three products shown in Figure 5 (c). Each of the products is put down and picked up 14 times on the shelf at three different locations marked as orange crosses. The environment is noisy with ambient acoustic noises such as foot stepping and speech. We also record the ground truth video.

Event Detection. We set the median filter window sizes as one second and the moving average filter window size as 0.5 seconds.

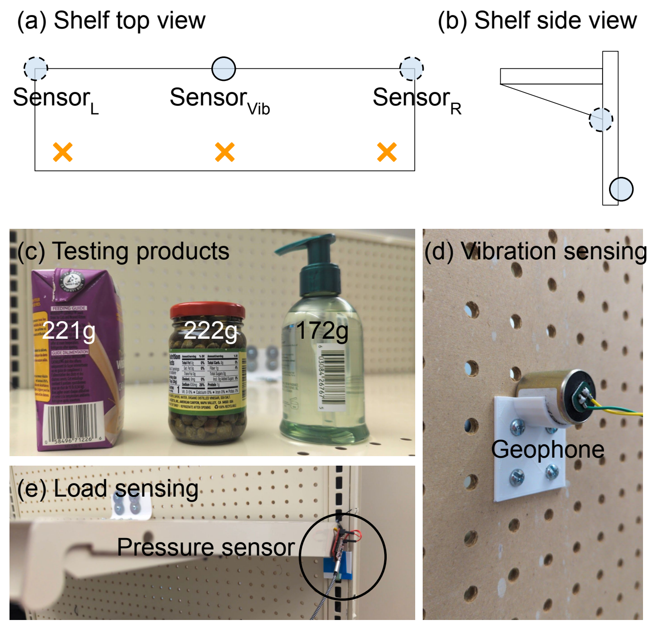

Figure 5: Experimental setup. (a) and (b) shows the sensor placement relative to the shelf with circle marks. The dashed line circles represent pressure sensors, and the solid line circles represent the vibration sensor. (c) lists the testing products with similar weights. (d) shows the vibration sensing step with a geophone sensor placed on the back panel. (e) depicts load sensing setup with the pressure sensor placed at the output arm of the right angle lever.

For event detection, we empirically set the threshold Th_σ as 0.5, and Th_{event} as 0.2 seconds. **Baseline:** We consider the event detection of vibration sensing described in Section 3.3 as the baseline.

Product Recognition. We set the threshold Th_{align} as 0.2 seconds. From the exploratory signal analysis, we observe that the frequency bands between 10-90 Hz and 300-400 Hz are the most representative, thus we select frequency components in these bands as features. We repeat the product recognition experiment 100 times and report the average F1 score and its standard deviation. In each repetition, we conduct a stratified sampling to randomly sample 80% of the data as the training data, and the rest forms the testing data. The stratified sampling process empirically guarantees the distribution of each product's vibration signal segments is consistent between training and testing data. The hyperparameter tuning process is conducted with a 5-fold cross-validation grid search in the training data. We adopt the same sets of candidate values (0.001,0.01,0.1,1,10) for the "C" and "γ" parameters. Finally, we test the model in the testing data with the selected best hyperparameter setting. **Baseline:** We also conduct the load sensing only product recognition as the baseline. The absolute value of difference in the windows of $[T_i{-}0.6s, T_i{-}0.2s]$ and $[T_i{+}0.2s, T_i{+}0.6s]$ plus the σ for the i^{th} event are used as features for recognition, and the SVM model's setting is identical.

Table 1: Load Distribution Analysis

Load Location	$Sensor_L$ Range (mV)	$Sensor_R$ Range (mV)
left	36.9(1.85)	–
middle	18.2(3.04)	27.3(2.71)
right	–	48.7(3.13)

Table 2: Additional Load Analysis

Additional Load	$Sensor_L$ Range (mV)	$Sensor_R$ Range (mV)
0	36.9(1.85)	48.7(3.13)
2 Kg	30.9(3.54)	41.6(2.50)
4 Kg	21.2(1.93)	26.5(2.84)

Evaluation Metrics. We use the F1 score as the event detection and product recognition performance metric, calculated as:

$$\mathrm{F1} = \frac{\mathrm{TP}}{\mathrm{TP} + \frac{1}{2}(\mathrm{FP} + \mathrm{FN})}, \quad (1)$$

where TP is number of true positives, FP is number of false positives, and FN is number of false negatives. For event detection, if the interval between the detected event and ground truth (i.e., labeled from video) is less than 0.2 seconds, we report it as TP, otherwise, it is FP. For production recognition, if the predicted production ID is match with the ground truth (interacted product ID), we report it as TP.

4.2 Results and Analysis

We analyze both the low-cost load sensing approach and the multi-modal task-level fusion performance.

4.2.1 Pressure Sensor Characterization. We use a soda can (380 grams) as the standard load change on the shelf and pick up/put down it at different locations (left, middle, and right) as shown in Figure 5 (a). We also investigate the impact of addition load (0, 2 Kg, and 4 Kg) on the load sensing. We characterize the pressure sensors' output range by measuring the voltage change in response to each interaction event.

First, we demonstrate $Sensor_L$ and $Sensor_R$'s output difference in Table 1. We observe that the interaction event location impact the pressure sensor's output range – the closer the location is to the sensor, the larger the output changes (higher SNR). This is because when the load is applied at different locations on the shelf, the load distributions on the peg hook on two sides of the shelf are different. For example, when the load is applied on the left side of the shelf, it is mainly distributed on the left side peg hook and cause a detectable pressure change. As a result, the pressure changes on the right side peg hook is limited and almost not detectable. Therefore, we did not report the sensor output range in these cases.

Since the pressure sensor's sensitivity decreases when the pressure increases [1], we further characterize our load sensing scheme with different amounts of additional loads on the shelf. Table 2 shows the range of the sensor output changes caused by the pick-up/putting down of a soda can when different amounts of additional loads are placed on the shelf. We can observe that with the increase of the additional load, the sensor reading changes decrease.

4.2.2 Event Detection and Product Recognition. Figure 6 (a) presents the F1 scores of event detection using vibration and load sensing. The green bar represents the F1 score achieved by the vibration-based detection (0.38). The lower performance could be caused by the ambient acoustic noises, which contribute to false positive detection. The yellow bar depicts the F1 score of load sensing (0.98), which shows a significant advantage over the vibration-based approach in terms of on-shelf event detection (2.6×). Figure 6 (b) shows the recognition F1 scores, with the vibration sensing results depicted by the green bar (0.81) and the load sensing results by the yellow bar (0.50). The vibration-based approach outperforms the load sensing approach (1.6×), as the vibration signals contain more distinctive information about the contact between the product and the shelf. In summary, this result verified our task-level multimodal fusion design – load sensing for on-shelf interaction detection and vibration sensing for on-shelf interaction recognition.

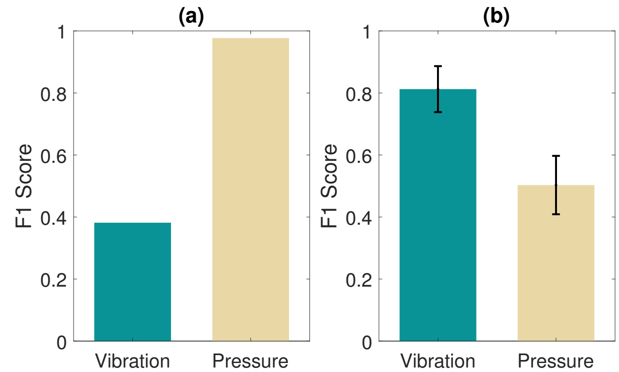

Figure 6: Results of customer-product interaction. (a) event detection accuracy. (b) product recognition accuracy.

5 DISCUSSION

The preliminary results are promising and lead to multiple technical directions we can explore.

Physical Augmentation for Load Sensing. In this study, we placed a pressure sensor directly between the retail shelf frame and the shelf. However, this placement could result in data drift over a long period of time when pressure is continuously applied to the sensor. In addition, the pressure sensor's sensitivity also varies in different pressure ranges. Therefore, it is important to dynamically calibrate the system for continuous and long-term monitoring. We plan to take the physical and data-driven combined approach, where we can modify both the mechanical placement of the sensor and the algorithm to reduce the impact of data drift and sensitivity variation.

Robust Inference Fusion. This preliminary work verifies the complementarity of the two low-cost modalities used for on-shelf product interaction recognition with a task-level fusion. We will further look into multimodal inference fusion techniques to further improve the accuracy and robustness of the interaction detection and recognition. For example, load sensing is efficient for distinguishing products with obvious weight differences, which can help the interaction recognition. In addition, when multiple products are picked up and put down, the combination of the load and vibration sensing can capture better spaciotemporal characteristics for recognition over mixed signals.

Low Power Design. Power and communication are two important aspects for us to further optimize the per-shelf system design for a scalable system, such as autonomous retail (hundreds of thousands of shelves). We will investigate and characterize the power consumption for each sensing modality. We plan to explore the low-power design and design trigger-based scheme for signal processing, communication, and inference.

6 CONCLUSION

In this paper, we present a multimodal solution that combines vibration- and load- sensing for customer-product interaction recognition in autonomous retails. We present a low-cost shelf load sensing scheme with sparsely deployed pressure sensors and introduce the multimodal sensing design that leverages the advantage of these two modalities. The load sensing is robust to ambient acoustic noise, while the vibration sensing achieves higher accuracy in recognition. We conduct real-world data collection with the proposed system, and our multimodal approach achieves 2.6× and 1.6× F1 score compared to the baseline method.

ACKNOWLEDGEMENT

This project was supported in part by AiFi Inc.

REFERENCES

[1] 2023. Force Sensitive Resistor. https://www.sparkfun.com/products/9376

[2] 2023. Geophone - SM-24 - SEN-11744 - SparkFun Electronics. https://www.sparkfun.com/products/11744

[3] AiFi. 2022. AIFI opens 80 computer vision-powered autonomous stores, an industry benchmark for platform scalability. https://www.prnewswire.com/news-releases/aifi-opens-80-computer-vision-powered-autonomous-stores-an-industry-benchmark-for-platform-scalability-301603023.html

[4] Yuanqiang Cai, Longyin Wen, Libo Zhang, Dawei Du, and Weiqiang Wang. 2021. Rethinking object detection in retail stores. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 947–954.

[5] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. 2019. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2019).

[6] Jonathon Fagert, Amelie Bonde, Sruti Srinidhi, Sarah Hamilton, Pei Zhang, and Hae Young Noh. 2022. Clean vibes: Hand washing monitoring using structural vibration sensing. *ACM Transactions on Computing for Healthcare (HEALTH)* 3, 3 (2022), 1–25.

[7] Howard Huang. 2021. Amazon go store review, Impressions, and thoughts. https://www.whatshuang.com/posts/amazon-go-store-review-impressions-and-thoughts

[8] Ji-Ye Jeon, Shin-Woo Kang, Hyuk-Jae Lee, and Jin-Sung Kim. 2022. A Retail Object Classification Method Using Multiple Cameras for Vision-Based Unmanned Kiosks. *IEEE Sensors Journal* 22, 22 (2022), 22200–22209.

[9] Murao Kazuya, Imai Junna, Terada Tsutomu, and Tsukamoto Masahiko. 2017. Activity Recognition and User Identification based on Tabletop Activities with Load Cells. 58, 1 (2017).

[10] Ming-Hong Lin, Muhammad Atif Sarwar, Yousef-Awwad Daraghmi, and Tsì-Uí İk. 2022. On-shelf load cell calibration for positioning and weighing assisted by activity detection: Smart store scenario. *IEEE Sensors Journal* 22, 4 (2022), 3455–3463.

[11] Lizheng Liu, Bo Zhou, Zhuo Zou, Shih-Ching Yeh, and Lirong Zheng. 2018. A Smart Unstaffed Retail Shop Based on Artificial Intelligence and IoT. In *2018 IEEE 23rd International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*. 1–4. https://doi.org/10.1109/CAMAD.2018.8514988

[12] Xiaochen Liu, Yurong Jiang, Kyu-Han Kim, and Ramesh Govindan. 2020. Grab: Fast and accurate sensor processing for cashier-free shopping. *arXiv preprint arXiv:2001.01033* (2020).

[13] Kazuya Murao, Junna Imai, Tsutomu Terada, and Masahiko Tsukamoto. 2015. Recognizing activities and identifying users based on tabletop activities with load cells. In *Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services*. 1–6.

[14] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[15] Jingtian Peng, Chang Xiao, and Yifan Li. 2021. RP2K: A Large-Scale Retail Product Dataset for Fine-Grained Image Classification. arXiv:2006.12634 [cs.CV]

[16] Yacine Rekik, Evren Sahin, and Yves Dallery. 2009. Inventory inaccuracy in retail stores due to theft: An analysis of the benefits of RFID. *International Journal of Production Economics* 118, 1 (2009), 189–198.

[17] George Roussos. 2006. Enabling RFID in retail. *Computer* 39, 3 (2006), 25–30.

[18] Carlos Ruiz, Joao Falcao, Shijia Pan, Hae Young Noh, and Pei Zhang. 2019. Aim3s: Autonomous inventory monitoring through multi-modal sensing for cashier-less convenience stores. In *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. 135–144.

[19] Vladimir Vapnik. 1999. *The nature of statistical learning theory*. Springer science & business media.

[20] Yue Zhang, Carlos Ruiz, Shubham Rohal, and Shijia Pan. 2023. CPA: Cyber-Physical Augmentation for Vibration Sensing in Autonomous Retails. In *Proceedings of the 24th International Workshop on Mobile Computing Systems and Applications*. 8–14.