


RESEARCH ARTICLE

Context-aware crowd monitoring for sports games using crowd-induced floor vibrations

Yiwen Dong¹ , Yuyan Wu¹, Yen-Cheng Chang², Jatin Aggarwal¹, Jesse R Codling², Hugo Latapie³, Pei Zhang² and Hae Young Noh¹

¹Department of Civil and Environmental Engineering, Stanford University, Stanford, CA, USA

²Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA

³Cisco Research, Cisco Systems, Inc., San Jose, CA, USA

Corresponding author: Yiwen Dong; Email: ywdong@stanford.edu

Received: 09 April 2024; **Revised:** 21 October 2024; **Accepted:** 12 June 2024

Keywords: context-aware; crowd monitoring; floor vibration; sports

Abstract

Crowd monitoring for sports games is important to improve public safety, game experience, and venue management. Recent crowd-crushing incidents (e.g., the Kanjuruhan Stadium disaster) have caused 100+ deaths, calling for advancements in crowd-monitoring methods. Existing monitoring approaches include manual observation, wearables, video-, audio-, and WiFi-based sensing. However, few meet the practical needs due to their limitations in cost, privacy protection, and accuracy.

In this paper, we introduce a novel crowd monitoring method that leverages floor vibrations to infer crowd reactions (e.g., clapping) and traffic (i.e., the number of people entering) in sports stadiums. Our method allows continuous crowd monitoring in a privacy-friendly and cost-effective way. Unlike monitoring one person, crowd monitoring involves a large population, leading to high uncertainty in the vibration data. To overcome the challenge, we bring in the context of crowd behaviors, including (1) *temporal context* to inform crowd reactions to the highlights of the game and (2) *spatial context* to inform crowd traffic in relation to the facility layouts. We deployed our system at Stanford Maples Pavilion and Michigan Stadium for real-world evaluation, which shows a 14.7% and 12.5% error reduction compared to the baseline methods without the context information.

Impact Statement

This paper introduces a novel approach for crowd monitoring using crowd-induced floor vibrations. It overcomes the long-standing limitations in the existing systems by enabling low-cost, non-intrusive, and continuous crowd monitoring in sports stadiums. However, one unique challenge of vibration-based crowd monitoring is the high uncertainty in data due to the indirect observation of crowd behaviors through the floor. To address this issue, we bring in the context of the game highlights and the stadium layout to reduce uncertainty by developing a probabilistic crowd-context association model. We evaluated this approach at two real-world stadiums and achieved promising accuracy. The development of this approach breaks the practical barrier in crowd monitoring, which has the potential to be widely adopted in sports stadiums, concert halls, auditoriums, and so on.

1. Introduction

Crowd monitoring involves tracking and analyzing the behavior of large groups of people during large-scale public events, such as sports games (Lamba and Nain, 2017). In sports stadiums, crowd monitoring is critical to ensure public safety (Zeitl et al., 2009), enhance the game experience (Filingeri et al., 2017), and optimize resource allocation (Molloy et al., 2009; Kingshott, 2014). Over the years, mismanagement of the crowd has led to grave consequences. For example, the Kanjuruhan Stadium disaster and the Seoul Halloween Stampede have caused 135 and 159 deaths, respectively (Sharma et al., 2023; Yogadhita and Agustin, 2023). Nevertheless, studies have found that such incidents can be prevented by proper crowd monitoring and timely crowd control (de Almeida and von Schreeb, 2019). By analyzing the behaviors of the crowd, we can detect and prevent potential threats, such as stampedes and violence (Wang, 2021). Moreover, we can also gain insights into the social and psychological aspects of crowd behavior, such as audience emotions and motivations (Zhang et al., 2018).

While many existing approaches are developed to monitor crowd behaviors, few meet the practical needs due to the high cost, privacy concerns, and reliability (Andersson et al., 2009; Kantarci and Mouftah, 2014; Al-Shaery et al., 2020; Haque et al., 2020; Kumar et al., 2021). For example, manual monitoring is the most common approach (Kok et al., 2016), which is interpretable and reliable when the staff are well-trained. However, it is labor-intensive, costly, and may have negligence and miscommunication issues. On the other hand, sensing devices such as cameras and microphones can significantly reduce the labor requirement and are more reliable, but they suffer from privacy issues due to the video and voice recordings of people in public spaces (Maheshwari and Heda, 2016; Bahmanyar et al., 2019). To reduce such concerns, WiFi- and radio frequency-based devices are used to capture the body motion of individuals without showing people's faces (Yamin et al., 2018; Jarvis et al., 2019), but they have difficulty in capturing the activity among a large group of people due to significant noise interference, producing inaccurate results.

In this paper, we introduce a novel method for crowd monitoring using floor vibrations. By placing vibration sensors underneath the bleachers or on the floor surfaces, we capture crowd-induced vibrations to infer crowd behaviors in terms of (1) crowd reactions (e.g., clapping, stomping, dancing) and (2) crowd traffic (i.e., the number of people entering each door). The primary insight is that various types of crowd behaviors induce distinct vibration patterns of the floor, allowing crowd behavior characterization and inference using the vibration signals. The main benefits of using floor vibration are that it is cost-efficient, non-intrusive, allows continuous monitoring, and is perceived as more privacy-friendly than cameras or audio recordings. This sensing approach has been explored in many existing applications, such as occupant detection (Reuland et al., 2017; Mirshekari et al., 2020), identification (Pan et al., 2017; Dong et al., 2023b), activity recognition (Alwan et al., 2006; Pan et al., 2019), localization (Mirshekari et al., 2018), and in-home health monitoring (Kessler et al., 2019; Dong and Noh, 2024; Dong et al., 2024a, 2024b).

However, the main challenge in vibration-based crowd monitoring is the high uncertainty in crowd behaviors combined with the indirect measurements made by the vibration sensors. Unlike monitoring a single person, crowd monitoring typically involves a large group of people (typically more than 1,000), which means the uncertainty in behavior patterns is significantly higher. Specifically, there are huge variations in the uniformity of their behaviors, particularly reflected in two aspects (1) During the game, floor vibration induced by crowd reaction is uncertain due to the difference among individuals and the proportion of people reacting; (2) Before/after the game, floor vibration induced by crowd traffic is uncertain due to the large range of a possible number of pedestrians. On top of that, unlike cameras, vibration sensors do not directly observe the crowd's behavior, making it difficult to examine the uncertainty. As a result, estimating crowd behaviors using the vibration signals may lead to larger errors than our prior vibration-based monitoring of a single person.

To overcome the high uncertainty challenge, we bring in the context of the game to reduce the estimation error. Specifically, we incorporate (1) the *temporal context* between the crowd reaction and the game progress, such as clapping after the home team's scoring, stomping/booing to disturb the opponents'

attack, and (2) the *spatial context* between the crowd traffic and facility layouts, such as people accumulating around the doors with food stands or restrooms. We first establish the association between the context and the vibration signals to make our data context-aware. Then, we formulate the crowd monitoring problem through a probabilistic graphical model to describe the relationship between the crowd behavior, vibration data, and context. Through this model, our method first learns the latent representations of the context (i.e., the game progress and facility layouts) through a neural network encoder, and then integrates the heterogeneous context information and vibration data to infer crowd behaviors. With context-data association modeling, our method mitigates the estimation error due to the uncertainty of a large of people, leading to more accurate and interpretable crowd monitoring.

The key contributions of this paper are:

- We introduce a novel crowd-monitoring method using crowd-induced floor vibrations, which continuously monitors crowd behaviors including crowd reactions and crowd traffic in sports stadiums.
- We model the relationship between vibration data, crowd behaviors, and spatial/temporal context to develop a context-aware probabilistic model for crowd behavior inference, reducing the uncertainty in vibration-based crowd monitoring data.
- We evaluate our approach through real-world deployments at Stanford Maples Pavilion and Michigan Stadium to assess its effectiveness and robustness under various scenarios.

Through two real-world deployments, our results show a 0.89 F-1 score in crowd reaction monitoring and nine mean absolute error (MAE) in crowd traffic (i.e., headcount) estimation among various entry doors, which has a 14.7% and 12.5% error reduction compared to the baseline methods without context information, respectively.

For the rest of this paper, we first characterize crowd-induced floor vibrations for crowd behavior monitoring (Section 2). Then, we formulate the context-aware crowd-monitoring model and introduce the detailed data processing process of our approach (Section 3). Next, we describe our real-world evaluation and discuss the results (Section 4). After summarizing the related work (Section 5), we conclude the study and present the future work (Section 6).

2. Characterization of crowd-induced floor vibration

In this section, we characterize the vibration signals to understand how crowd behaviors are inferred from the data. We first discuss its relationship with the crowd reaction and then the crowd traffic.

2.1. Relationship between crowd reaction and floor vibration

We characterize the floor vibrations induced by various types of crowd reactions, including (1) quiet (no body motion), (2) active (sitting with upper or lower body movements such as clapping and foot shuffling), and (3) moving (standing/walking with lower body movements such as stomping and dancing), as shown in Figure 1. The vibration induced by the quiet reaction has a low signal amplitude and noise-like oscillations around the mean. In contrast, the vibration induced by moving (i.e., stomping) has large amplitudes, characterized by separated impulses each representing a heavy step. Other active reactions such as clapping induce floor vibration indirectly through the bleacher seats, so the signal has a lower amplitude than moving with a unique frequency representing the physical properties of the seat-floor connection.

2.2. Relationship between crowd traffic and floor vibration

The number of audience entering through each door can be inferred from floor vibration signals, captured by vibration sensors placed beside each door on the surface of the floor. The insight behind this method is

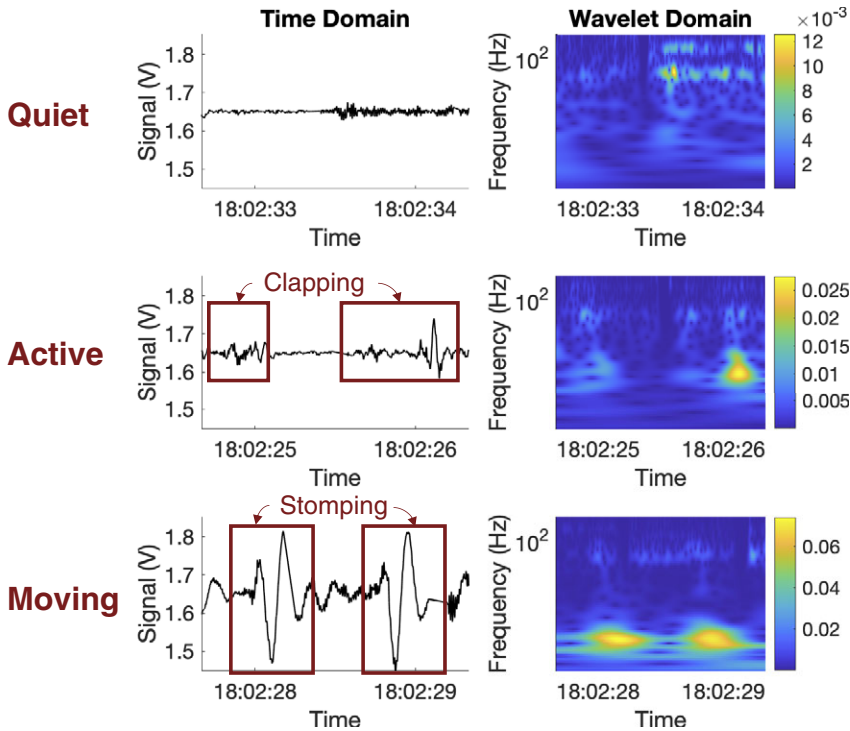


Figure 1. Characterization of vibration signals induced by crowd reactions, including quiet, active, and moving (from top to bottom). Both time- and wavelet-domain plots show clear distinctions among various crowd reaction types.

that audience movements, such as walking, opening, and closing doors, induce the floor vibrations. These vibrations are detected by the vibration sensors (as illustrated in Figure 2). It is observed that there is a direct correlation between the major peaks' intervals within the vibration signals and the number of individuals passing through the doors. Therefore, these peak intervals within a certain time suggest the frequency of audience movements, including walking and door operations within this time. For instance, increased footstep and door operation events correspond to higher crowd traffic. Consequently, identifying the number of peaks in the vibration signals, which are related to the audience movement, facilitates the estimation of crowd traffic at each entry point.

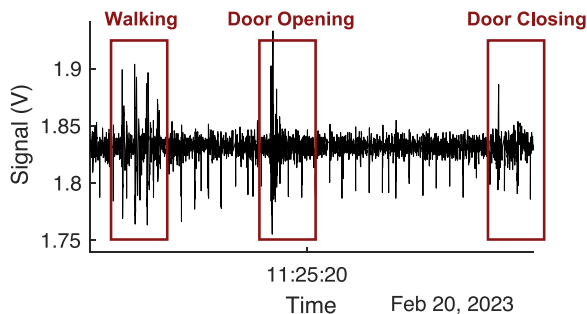


Figure 2. Characterization of vibration signals induced by crowd traffic near an entry door. The level of crowd traffic can be indirectly inferred from the peaks detected in the signals induced by walking and door opening/closing.

3. Context-aware crowd monitoring through crowd-induced floor vibrations

In this section, we first provide an overview of our approach with an emphasis on the formulation of a context-aware probabilistic model, and then describe each component in detail to illustrate the data processing and modeling procedures.

3.1. Overview

The main idea of our approach is to integrate the crowd-induced vibration data with the spatial and temporal contexts through neural network encoders to collectively infer crowd behaviors in terms of crowd reaction and traffic, as summarized in Figure 3.

First, crowd-induced vibration data are collected by sensors mounted underneath the bleachers or attached to the floor surface near the entry doors. The vibration signals are transmitted wirelessly to a centralized server and stored in a hard drive. We preprocess the raw signals through filtering, sliding windows, and interpolation algorithms to produce discretized signal segments for analysis.

Then, we associate the contexts and vibration data through temporal and spatial relationships with game progress and the facility layout, respectively. The contexts are provided by the game organizer/venue manager, which includes the progress of the game with detailed scores and section breaks over time, as well as the layout of facilities (e.g., the location of restrooms and food stands) around the venue. The vibration data and the contexts are associated based on the temporal and spatial proximity, which will be discussed in Section 3.3.

After that, we model the contexts and the vibration data to make inferences. They are first processed through a context encoder and a vibe data encoder, respectively, to extract representative features. Then, we integrate the context with the processed vibration data through an update encoder to infer crowd behaviors. The output of our approach includes (1) crowd reaction, such as the crowd status (e.g., quiet, active, and moving) and the specific type of reaction under each status (e.g., clapping and stomping), and (2) crowd traffic in terms of the number of people entering each door.

3.2. Task-specific crowd behavior detection and feature extraction

In this section, we detect the crowd behaviors in the vibration data and extract task-specific features for subsequent data modeling. The data processing pipeline differs depending on the specific tasks, which are (1) crowd reaction and (2) crowd traffic monitoring.

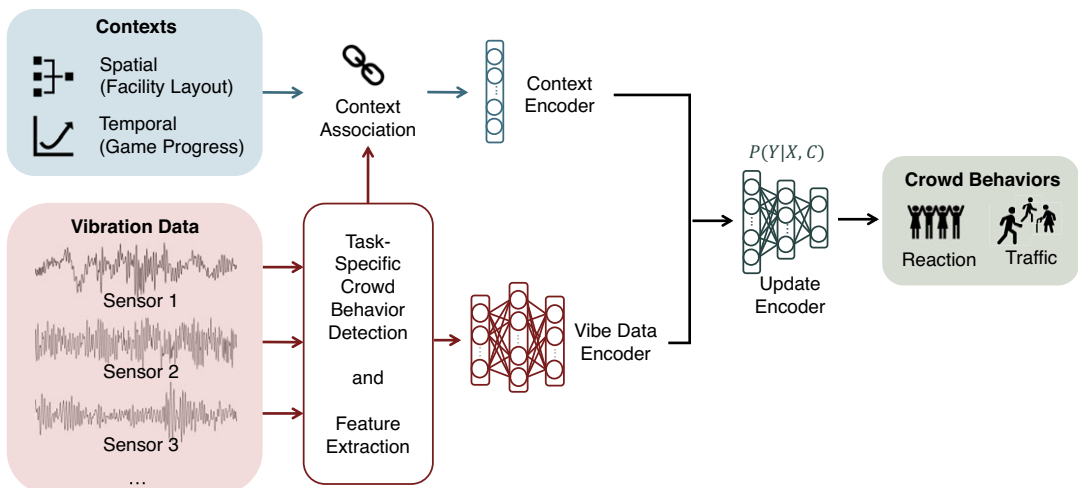


Figure 3. Our context-aware crowd monitoring approach integrates crowd-induced vibration data and spatial/temporal contexts to estimate crowd behaviors.

We first preprocess the signals to prepare for window-based crowd behavior detection and mitigate the missing data issue for subsequent data modeling. First, we segment the signals into 1-s windows with a time step of 0.5 s to avoid the effect of the activity signals truncated by the window edges. If more than half of the data is missing in a window, it is excluded from further analysis. Conversely, if the amount of missing data within a window is less than half of its duration, a linear interpolation method is employed to efficiently fill in the missing values.

For crowd reaction monitoring, we detect crowd behavior by identifying vibration windows that capture crowd reactions from the sensors underneath the bleachers. To detect these windows, we compare these signal windows with a reference noise signal, which is selected as the 2-min window during inactive periods (when the stadium is empty). These selected windows serve as our noise signal baseline. Assuming a normal distribution for the noise, we compute the mean and standard deviation for these noise signals for each sensor. We then adjust the signals by subtracting the mean value of the noise signal to ensure a zero average. Considering the noise variance during the game, which is due to the loud music (observed to be around 20 times the noise's standard deviation), we set 20 times the standard deviation of the noise signal as the threshold for detecting crowd reaction. Vibration signals that exceed this amplitude threshold are regarded as the time when audience activities happen. To improve the accuracy of our analysis, we apply a moving median smoothing over adjacent five windows to these detected activities. As demonstrated in Figure 4, this method successfully identifies 83% of crowd reactions within a 5-s error range of the labeled reactions.

For crowd traffic monitoring, we leverage the observation in Figure 2 to identify crowd behaviors related to the traffic by detecting the vibration signal peaks from sensors near the entry doors. We identify the peaks in the vibration signals by setting a minimum peak distance of 1 s and selecting a threshold as the minimum peak height. The threshold peak height for peak detection is selected based on the correlation score between the number of detected peaks and the actual count of people entering, which is determined through a preliminary analysis of the training data. The chosen threshold corresponds to the highest correlation score observed during this analysis. The detected peaks are used to extract features for crowd traffic estimation.

After the crowd behavior detection, we extract features from the vibration signals for crowd reaction and traffic monitoring, respectively.

Crowd reaction features. The vibration features for crowd reaction monitoring include:

- *Time-domain feature*: signal energy of each 0.1-s segment.
- *Frequency-domain feature*: sum of signal amplitudes in each 10 Hz range after the Fourier transform.
- *Time–frequency-domain feature*: sum of wavelet coefficients within each 10-Hz and 0.1-s grid block.

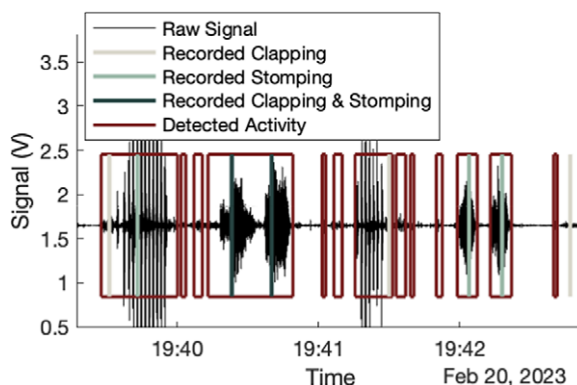


Figure 4. Crowd behavior detection results on sample vibration data. The algorithm successfully detects 83% of crowd reactions within 5-s error ranges of the ground truth records.

These features offer a more comprehensive depiction of crowd reactions, involving various dimensions of vibration signals, such as temporal variations, frequency characteristics, and the dependency between time and frequency domains. A random forest model is employed to rank the importance of these features. This importance is determined based on the impurity of crowd reaction types after splitting data on a feature. The more a feature decreases the impurity, the more important the feature is. This provides an efficient process to reduce the feature dimension while preserving the important information in vibration data.

Crowd traffic features. The vibration features for crowd traffic estimation include:

- *Peak count feature:* the number of peaks detected, which indicates the frequency of people interacting with the door and stepping on the area of the floor near the sensors.
- *Peak height features:* the maximum, minimum, average, and standard deviation of the height of the detected peaks, which describe the movement type (e.g., footsteps vs. door opening) and how urgent the movements are.
- *Peak time difference features:* the minimum, maximum, average, and standard deviation of the time differences between adjacent peaks, representing the movement frequencies.

These features are extracted based on our observation in [Figure 2](#) that the peaks of the floor vibration signal represent the crowd traffic patterns such as walking, door opening, or door closing.

To overcome the limited size of our dataset for crowd traffic estimation, we also augment our sample by merging two 10-min windows in the original sample to generate a new sample. This is based on the assumption that the relationship between the crowd traffic and floor vibration at each door is not affected by time (i.e., time-invariant) because we use the same sensor and put it at the same location throughout the game. The data augmentation is realized by generating a new sample by merging the features of the two windows. The output ground truth of each augmented sample is obtained by summing up the number of people entering within these two windows.

3.3. Establishing context associations

To incorporate the influence of contexts (i.e., game progress and facility layout) on crowd behaviors, we establish context associations to provide temporal and spatial contexts to the detected crowd behaviors in the vibration data. Specifically, we establish (1) the *temporal context association* between crowd reaction and the game progress ([Section 3.3.1](#)) and (2) the *spatial context association* between the crowd traffic and facility layout ([Section 3.3.2](#)). With these contexts, we formulate context-aware probabilistic models to allow more accurate and interpretable estimation of crowd behaviors ([Section 3.3.3](#)).

3.3.1. Temporal context association

The *temporal context association* is defined as the relationship between crowd-induced floor vibrations and the game progress through their occurrences in time sequence. For example, if the crowd reacts with a round of applause after a home team's goal, there is an association based on the time sequence such that "clapping" occurs concurrently or right after the "goal" event.

We establish the *temporal context associations* based on the time sequence of occurrences between events during the games and crowd-induced vibration signals. In the previous example, we capture the unique floor vibration signals induced by the "clapping" motion to establish an association with the "goal" event, which provides a context of the game to the vibration data recorded at that time. The game event types we focus on are mainly the score changes and game time divisions (i.e., playing periods and game breaks), which are easily accessible from the stadium operation team. The vibration signals are divided into a series of 1-s windows for discrete association. [Figure 5](#) shows a snapshot of the *temporal context associations* between game progress and vibration data. A series of continuous active signal windows (black dots) are matched with various events (colored dots) in the games. With the established association, we assign a unique embedding to each window to represent the effect of its corresponding context through

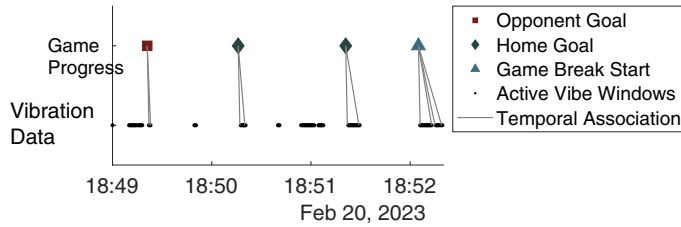


Figure 5. Temporal context association between crowd-induced vibrations and the game progress. Each game event such as the opponent goal, home team goal, and game break start is matched with active vibration windows over time.

the concatenation of encoded data features and context embeddings before feeding into the update encoder (see Figure 3). However, not all windows are matched because these vibrations may be induced by unrecorded game events, such as extraordinary blocking and passing moments, or by individuals who are sitting near the sensors. In these cases, we estimate the crowd reaction through vibration data only.

With the relationship between floor vibration and crowd reactions, we further analyze how audience reaction changes as the game progresses. Figure 6 shows the distribution of crowd reaction types associated with various game events, including home goal, opponent goal, and game break. We observe that the crowd is mainly clapping (i.e., active) after the home goal, while remaining quiet or booing after the opponent’s goal, except for a few opponents’ fans. The moving (mainly stomping) reaction observed at the opponent’s goal is mainly caused by noise making to distract the opponent during the defense, showing support for our home team. Crowd reactions during the game break are more diverse as people may choose to take a break by leaving the seating area or stay to enjoy the entertainment on-court such as dance performance, kids’ mini-game, T-shirt toss, and dance/kiss cams.

3.3.2. Spatial context association

The spatial context association refers to the relationship between the vibration data recorded at entry doors and the facilities around those doors based on spatial distance. For instance, when a sensor is positioned at an entrance door near a food stand, the vibration data will exhibit more peaks, indicating higher foot traffic compared to a door without such facilities. As depicted in Figure 7, door 1, which is located near more facilities shows a notably higher volume of people entering compared to door 5.

We establish the spatial context associations based on the distance between each sensor and various facility types, including restrooms, game swag stations, food stands, and the student center. The associations are encoded by a spatial proximity matrix (with rows representing the doors and columns representing the inverse proximity from each facility type) to enhance the accuracy of our crowd traffic

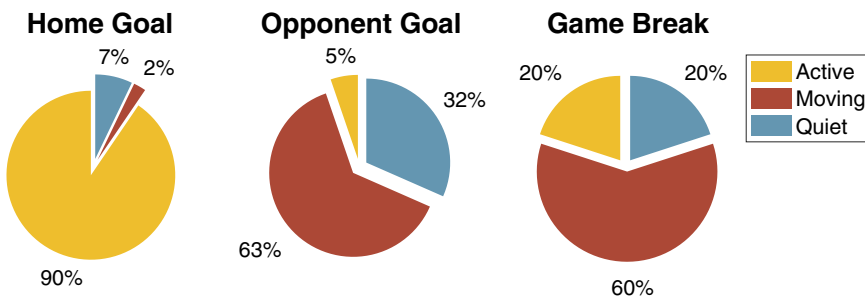


Figure 6. Crowd reaction distribution varies across various events in a sample game. Active (clapping) dominates the home goal event, while moving (stomping) and quiet dominate the opponent goal event. The reactions are more evenly distributed during the game break.

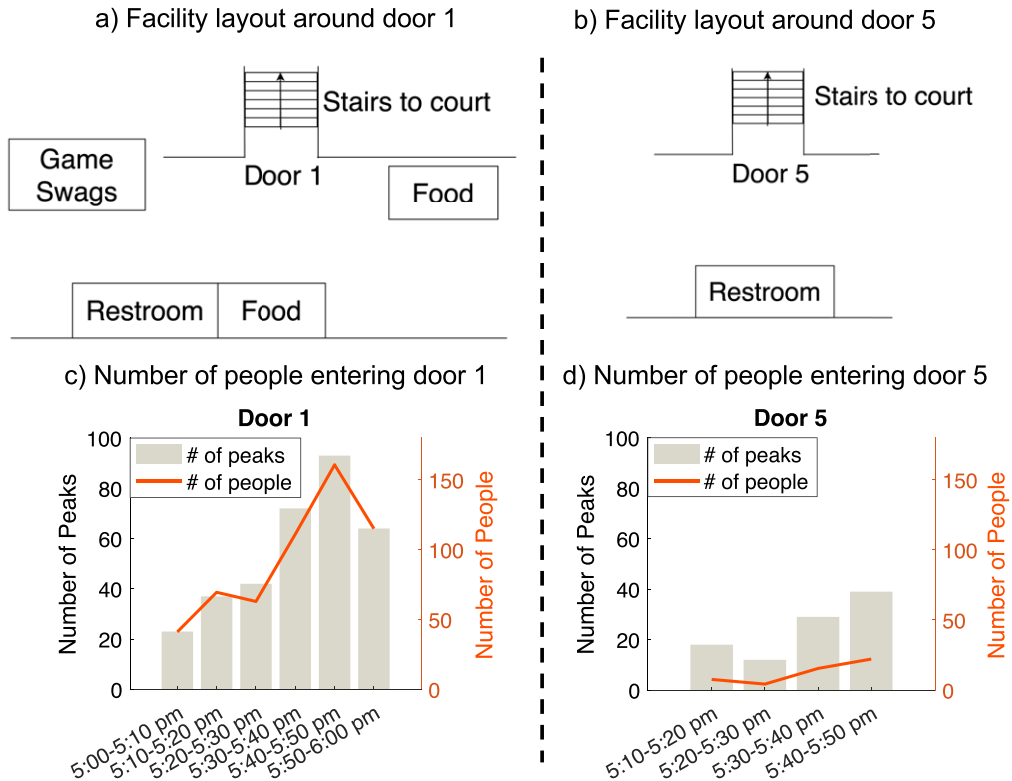


Figure 7. Spatial context association around entry doors. The difference in facility layout around (a) door 1 and (b) door 5 leads to distinct crowd traffic as shown in (c) and (d). The crowd traffic is correlated with the number of peaks detected in the vibration signals.

analysis, which has been introduced in our prior work (Dong et al., 2023c). First, the distances between the facilities and the doors are estimated based on the shortest path. Then, an overall maximum walking distance threshold is determined as the maximum distance to reach any facility starting from the closest door. With these two distances, the proximity is computed as the ratio between the overall maximum distance and the actual distance between the door and various facilities. The higher the ratio is, the closer the facility is to the door. This number is crucial because it determines the strength of contextual influence on crowd traffic: the shorter the distance is, the more significant the facility's influence is on crowd traffic.

Figure 7 shows that the layout of essential facilities such as food stations, restrooms, and game swags plays a crucial role in determining crowd traffic. For example, the facility layout around two sample doors at our evaluation site was obtained from the stadium map provided by the venue operator. Door 1 is surrounded by two food stands, a restroom, and a game swag station, which attracts a larger audience. In contrast, door 5 has fewer facilities around, which leads to a lower level of crowd traffic. As we compare Figure 7(c) and (d), however, we observe that the ratio of detected peaks to the actual crowd traffic differs across these two doors. This means that only knowing the number of peaks in the vibration signals is not sufficient to estimate the number of people. Therefore, we leverage the distinct layout of facilities surrounding door 1 and door 5 to enhance the accuracy of our crowd traffic estimation through facility associations, which is introduced in Section 3.3.3.

3.3.3. Developing context-aware probabilistic model using context associations

With the temporal context and spatial context associations established in the previous subsection, we formulate the crowd monitoring problem by developing a probabilistic model that formalizes the

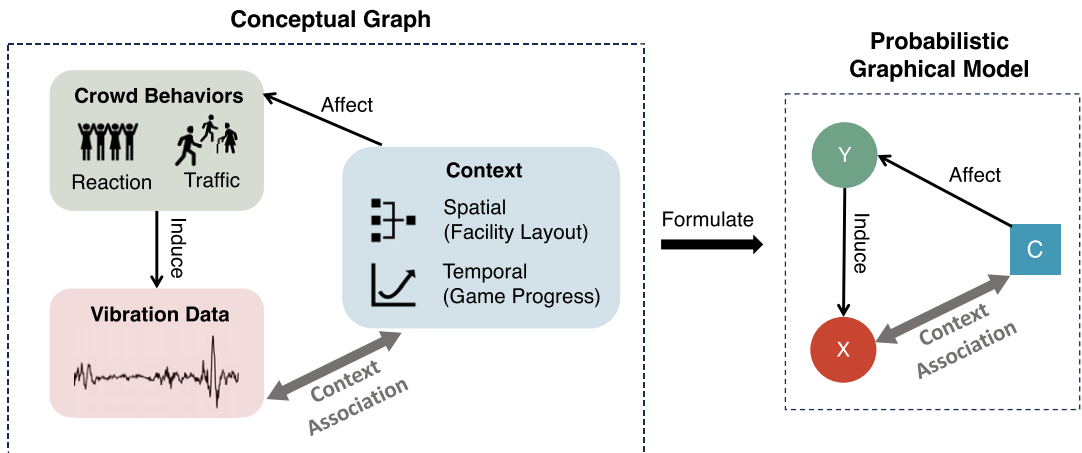


Figure 8. Conceptual graph (left) and the corresponding probabilistic game/facility association model (right). The graph describes the relationships among the crowd reaction (Y), spatial and temporal contexts (C), and vibration data (X) through context associations.

relationship among vibration data, contexts, and crowd behaviors. As summarized in Figure 8, the conceptual graphs (left) describing such relationships are converted into the corresponding probabilistic graphical models (right), allowing probabilistic analysis of the crowd behavior through vibration data.

Figure 8 shows the conceptual graph between crowd behaviors, vibration data, and spatial/temporal contexts. According to the discussion in Section 3.3, the game progress affects the crowd reaction, and the facility layout affects the crowd traffic. Both contexts can be associated with the vibration data induced by the crowd behaviors. To this end, we formulate a probabilistic graphical model among crowd behavior (Y), contexts (C), and vibration data (X) based on the conceptual graph, where dependencies and context associations are maintained. Assuming that the game record and facility layout are accurate, C is regarded as a deterministic variable in our model. With this formulation, the objective of crowd monitoring is to estimate $P(Y|X,C)$.

3.4. Modeling and integrating context and vibration data

In order to estimate the crowd behaviors given the data and contexts (i.e., $P(Y|X,C)$) formulated in Figure 8, we develop task-specific models to achieve crowd reaction and traffic monitoring, respectively. Specifically, the crowd reaction model is trained using the sensors underneath the bleachers integrated with the temporal context, while the traffic model is trained using the sensors mounted on the floor surface near the entry doors integrated with the spatial context.

3.4.1. Temporal context-aware crowd reaction monitoring

To model the temporal context (i.e., game progress) and bleacher vibration data, we design two neural networks with different characteristics to encode the features. First, we leverage a context encoder to learn the latent variables representing the multifaceted influence of the game progress on crowd-induced vibrations (e.g., intensity and duration) by expanding a one-dimensional score change or game break indicator to a multidimensional vector. Then, we use a vibe data encoder to learn latent representations of the crowd reactions from vibration data. The game record encoder is designed as a one-layer neural network that converts each 1-d game event into a 32-d vector, describing the process in which a single game event has multiple aspects of influence on the vibration data. The vibe data encoder is designed as a three-layer, 256-neuron wide neural network considering the complex inter-dependency between various selected features requires a larger number of neurons to capture. In addition, a 40% dropout is applied to

the neural network to mitigate the overfitting problem. The percentage of dropouts is chosen based on the performance during preliminary testing on data from one game.

After modeling the temporal context and vibration data, we concatenate their learned embeddings and integrate this information through an update encoder to estimate the conditional distribution $P(Y|X, C_t)$ as introduced in Section 3.3.3. The encoder is a three-layer funnel-shaped neural network that gradually transforms the concatenated embeddings to approximate the distribution of $P(Y|X, C_t)$. The resultant conditional probability is represented as a vector with the same length as the number of crowd reaction types.

3.4.2. Spatial context-aware crowd traffic monitoring

To model the spatial context (i.e., facility layout), we develop a spatial proximity matrix to encode its relationship with the vibration data. The spatial proximity matrix is a look-up table of the weight of each facility type (represented as rows) corresponding to the sensor at each door (represented as columns). The weight of each facility is determined by the ratio between the maximum walking distance to any facility from the closest door (around 20 m in our case) and the actual distance between the facility and the door. A higher weight means a shorter distance from the door to the facility, indicating a stronger spatial association. These facilities include food stands, game swag stations, restrooms, and game courts, as described in Figure 10. On the other hand, the vibe data encoder is a two-layer neural network with 32 neurons due to the smaller feature dimension than the previous module, which learns the latent variables of the crowd traffic from the vibration data.

With the vibration data and facility layout modeled at each door, we match the vibration data with its door and concatenate the learned embeddings from neural networks for spatial context-aware updating. The update encoder is a two-layer neural network that approximates the distribution of $P(Y|X, C_s)$. The output is the number of people entering each door.

4. Real-world evaluation in sports stadiums

We evaluated our approach at two real-world sports stadiums (1) Stanford Maples Pavilion and (2) Michigan Stadium, and obtained promising results in crowd reaction and traffic monitoring. In this section, we first introduce the deployment details at these two sites and then present the results of crowd monitoring. Furthermore, we discuss the variables that affect crowd behaviors and results, such as game types, sensor locations, promotional events, and crowd levels, suggesting future improvements.

4.1. Vibration sensing platform setups

The vibration sensing platform consists of robust, independent geophone sensing nodes that communicate over a private WiFi network in the stadiums. The geophones convert the vertical velocity of the floor into electrical signals which are digitized at the node and then transmitted to a centralized aggregator. At the aggregator, each geophone's data is recorded for later processing. These data can be analyzed at the aggregator or downloaded for analysis on a more powerful machine such as a laptop.

The setting of the large-scale sports stadium requires additional adaptations to our previous sensor network design that have been successful for monitoring on smaller scales (Bonde et al., 2021; Dong et al., 2023a). Many of the assumptions made in our previous sensing network no longer apply. For example, the mains-powered nodes were changed to battery-powered geophone sensors to avoid tripping hazards posed by wired charging in public spaces. In addition, the network topology is upgraded to a multi-hop mesh network as shown in Figure 9. In this mesh network, multiple wireless access points service connections to sensor nodes while passing data between each other on a separate wireless channel. The fully wireless nature of this setup was necessary not only for communication range but also to comply with strict visibility requirements and safety requirements to minimize tripping hazards for the public.



Figure 9. Sensor network design and multi-hop data transmission through routers via WiFi connections at the Stanford Maples Pavilion.

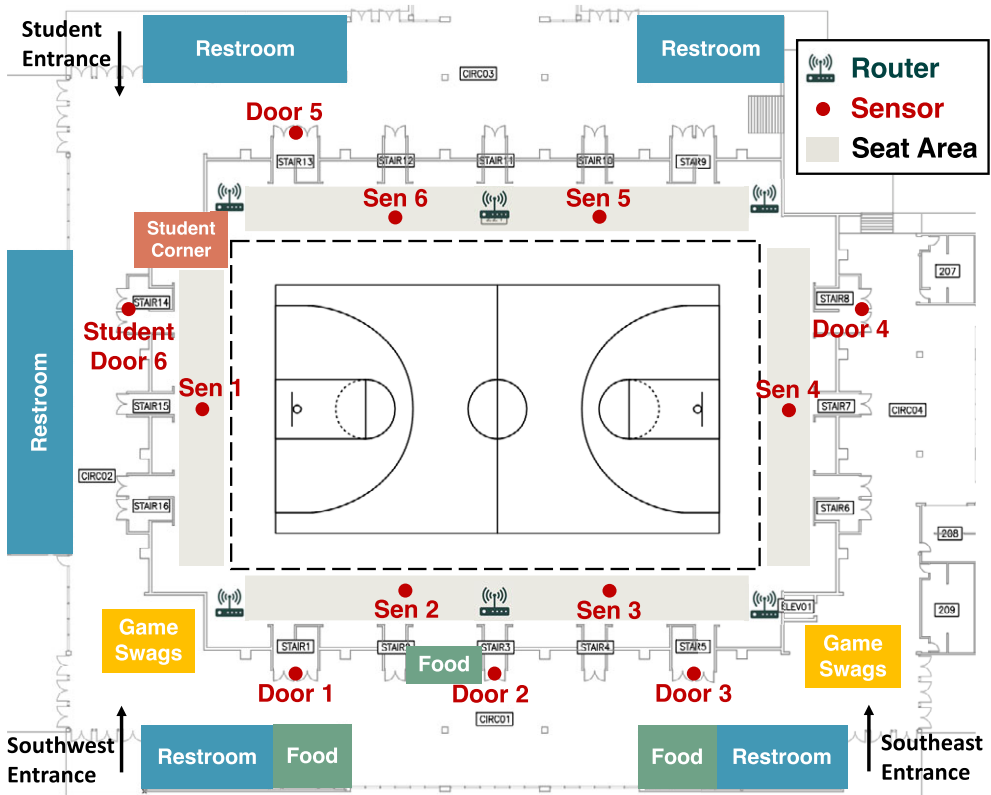


Figure 10. Experiment setup at Stanford Maples Pavilion with the sensor layout (marked as red dots), router layout (marked as green devices), facility locations at the concourse area (described as squares of different colors), and 16 entry doors connecting the concourse and the game court.

4.1.1. Deployment at Stanford Maples Pavilion

We conduct experiments for six NCAA women's and men's basketball games at Stanford Maples Pavilion, producing more than 280 h of vibration data from 12 sensors. The deployment setup is shown in Figure 10, which involves two sets of sensors (1) six interior sensors (sen 1–6) located underneath the bleachers at the seating area, and (2) six exterior sensors (door 1–6) located on the floor of selected entry doors connecting the concourse and the game court. The sensors are installed and uninstalled before and after each game for battery change and functional checking. All sensors are connected over a wireless mesh through six routers distributed across the venue as discussed in Section 4.1. The sampling frequency is set to 500 Hz to maximize the temporal resolution while ensuring data transmission efficiency.

The ground truths and the context information are collected through multiple sources, including (1) a volunteer team of 6–8 people observing the crowd for each game, (2) the ESPN website for score change over time (temporal context), and (3) the stadium management team for the facility layout (spatial context). Before and after the game, the volunteers count the number of people passing the doors with deployed sensors every 10 min. During the game, the volunteers are spread across the seating areas and record the crowd reactions around each interior sensor. The labels include (1) crowd status—quiet, active, moving, and (2) crowd reactions—clapping, stomping, dancing, and walking. The volunteers also record the playing period and the game breaks.

4.1.2. Deployment at Michigan stadium

At the Michigan Stadium, we conducted experiments during two NCAA men's football games, specifically the game of Michigan Wolverines vs. Ohio State Buckeyes which attracted 110,615 people to attend (Figure 11). Renowned as the largest stadium in the United States and the third-largest globally, the Michigan Stadium provides an ideal real-world setting to test our approach. Due to the large scale of the Michigan Stadium (twenty times of the Stanford Pavilion), we chose to cover part of the stadium space for evaluation purposes and deploy six vibration sensors around the stadium (shown in Figure 11b)

For our sensor deployment, we focus on interior sensors (sen 0–5), strategically positioned beneath the cement bleachers near the audio sections at stadium (Figure 11b). The sensors are installed and uninstalled before and after each game to ensure optimal performance, including battery replacement and functional checks. To facilitate seamless data transmission, all sensors are interconnected via a wireless mesh network, leveraging 16 routers placed around the stadium. The number of routers needed depends on the coverage of each router and the size of the stadium. The location of the routers depends on the availability of power outlets. Similar to the setup at Stanford Maples Pavilion, we configure the sampling frequency to

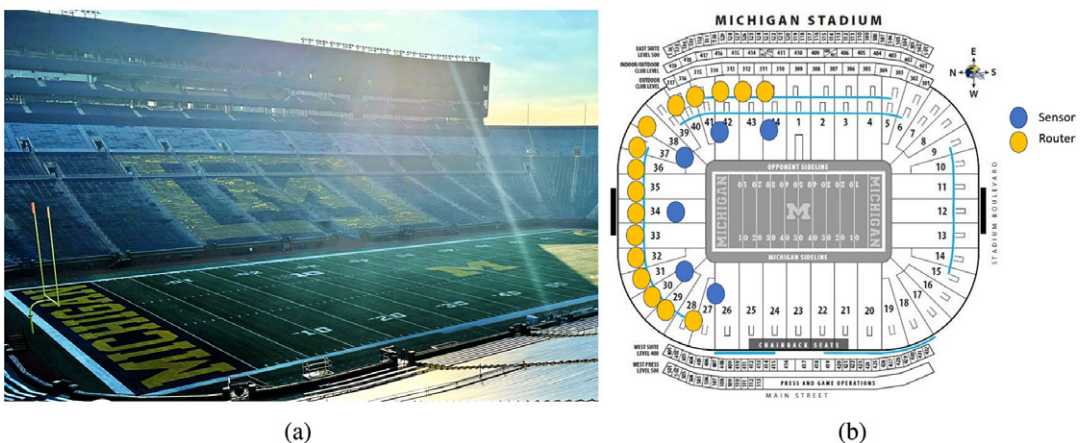


Figure 11. The experiment setup at Michigan Stadium includes the sensor layout (marked as blue dots), and the router layout (marked as yellow dots). Sixteen routers were strategically positioned to provide stable connections for our sensors.

500 Hz. This aims to maximize temporal signal resolution while maintaining efficient data transmission capabilities, enabling us to capture nuanced vibration signal patterns within the stadium environment effectively.

We gather ground truths and contexts from various information sources. First, we utilize the play-by-play game records from the EPSN website combined with manual crowd reaction labeling during the game. Figure 12 illustrates an example of our manual labels with respect to the vibration data collected in the stadium. These labels are defined based on a general observation of crowd reactions during the game. For instance, positive crowd reactions are associated with events such as a home team down, complete pass, or touchdown, while negative reactions correspond to similar events for the visiting team. Additionally, real-time score updates from the EPSN website provide the temporal contexts of game progress. Our labels include crowd status indicators, including quiet, active, and moving, as well as a diverse range of crowd reactions such as cheering, booing, moving, and storming.

4.2. Overall performance of our approach

Overall, our approach has a 0.89 F-1 score in crowd reaction monitoring and a 9.3 mean absolute error (MAE) in estimating headcounts for crowd traffic estimation among various doors. Compared to the context-only and data-only baselines, our method has significant improvements in crowd monitoring performance with a higher F-1 score (see Figure 13) and lower error (see Figure 15). The results are obtained by randomly splitting the data by 80% train and 20% test for each game, repeated 10 times to compute the average performance. The performance increase over the baselines is mainly because our approach incorporates the temporal and spatial contexts through the associations with the vibration data. Since the game drives the crowd reactions and the facility layout directs the crowd traffic, these contexts provide reliable prior information for crowd monitoring.

4.2.1. Crowd reaction monitoring performance

Our approach has a 0.89 and 0.83 F-1 score in crowd status and reaction classification, respectively. The overall performance for all test data has a 14.7% and 11.1% improvement compared with the context-only and data-only baseline, respectively. Figure 13 (a) shows that our method outperforms both baseline methods for all sensors. The improvement indicates that the temporal context corrects misclassified samples due to the uncertainty in vibration data. This is because the latent representations learned from the

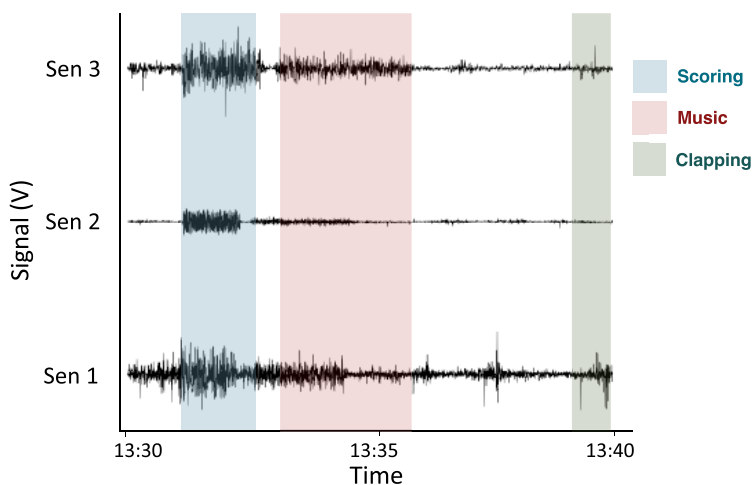


Figure 12. Visualization of vibration signals with respect to events happening during the game at Michigan Stadium. All sensors exhibit consistent trends when scoring, playing music, and clapping. The signal amplitudes induced by the same event vary across sensors due to different sensor locations.

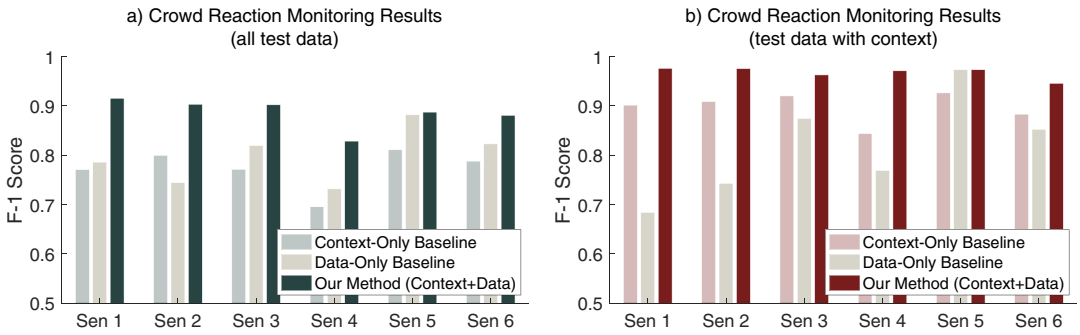


Figure 13. Crowd reaction monitoring results by comparing context-only, data-only baselines with our method. (a) Shows the results of all test data, where we observe that our method outperforms both baselines for all sensors. (b) Shows the results of data with contexts, where we observe that the context helps significantly with the monitoring results.

game progress can indirectly reflect the cause of crowd reaction variations, such as the active crowd size and the activity intensity. In addition, it is interesting to note that the F-1 score difference between context-only and data-only is not consistent among various sensors. This inconsistency suggests that sometimes context is more informative, while at other times, data is more informative, indicating their complementary nature. Figure 13 (b) shows that if the data has contexts, the contexts are typically more informative of the crowd reaction than the data itself. This means the crowd behavior may be mainly driven by the game progress when there are highlights on the court, while the rest of the crowd behaviors are more random and spontaneous so they need to be mainly inferred from the data.

One advantage of our vibration-based approach is that we provide fine-grained information on body movements (upper and lower body) through the signal patterns. Compared with wearables/mobile devices that only measure the motion of one body part or cameras that mainly look at the upper body, our vibration-based approach captures detailed movement characteristics through various frequency components. For example, the lower body movements directly exert forces on the floor (e.g., stomping and foot shuffling), inducing dominant frequencies of the floor. In contrast, the upper body movements indirectly induce vibrations through the chair (e.g., clapping and cheering), thus resulting in unique dominant frequencies from the chair–floor interactions.

Figure 14 shows the consistency between crowd reactions and game progress. Since the majority of the crowd at Michigan Stadium supports Michigan Wolverines, the crowd cheers when Michigan makes

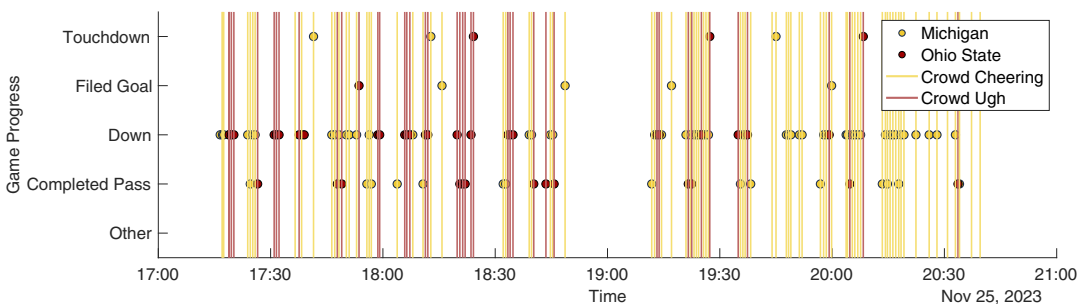


Figure 14. Visualization of the crowd reactions at critical game events over time. The crowd cheers (yellow vertical lines) when Michigan achieves completion, down, field goal, and a touchdown (yellow circles), while the crowd ughs (red vertical lines) when Ohio State makes progress (red circles). The alignment of crowd reactions and game progress justifies the efficacy of adding temporal context for crowd reaction monitoring.

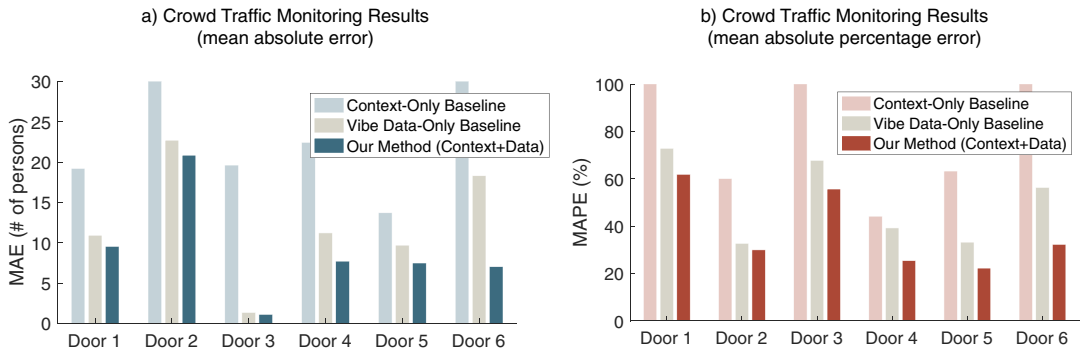


Figure 15. Crowd traffic monitoring results by comparing context-only, data-only baselines with our method. (a) Shows the mean absolute errors of our method are lower than both baselines for all entry doors. (b) Shows a relatively high error rate by percentage, indicating that vibration-based traffic monitoring is a challenging task and has room for improvements for future work.

progress (i.e., completed pass, down, field goal, and touchdown), which is visualized through the nearly perfect alignment between the yellow circles and the yellow lines. In contrast, the crowd expresses disappointment and “ugh” towards the moment when Ohio State Buckeyes make progress. The white gap in the middle indicates the game break after the first two quarters. This visualization shows the crowd activity frequency and intensity over time and justifies the effectiveness of the temporal context (i.e., game progress) in crowd reaction monitoring.

4.2.2. Crowd traffic estimation performance

Our approach has an average of 9.3 MAE for crowd traffic estimation, which has an average of 64.2% and 27.6% error reduction when compared to the context-only and data-only baseline methods. The context-only model has the highest error because it only considers the location of the entry door and thus remains fixed regardless of the game’s popularity and temporal variations. Nevertheless, the location information is still helpful for error reduction because it reduces the uncertainty in sensor data by associating the crowd traffic with the facility’s popularity around. We also compute the mean absolute percentage error (MAPE) of our method to the error rate, which is 37.9%, significantly lower than the context-only (206.5%) and data-only (50.4%) baseline methods. The relatively high percentage error means that vibration-based traffic monitoring is challenging and requires further study to improve the accuracy of the person counting. It is worth noting that MAPE explodes when a door has almost no one entering (which means the denominator is nearly zero), which often happens during the first 10 min of entry. In this case, a high error rate does not necessarily mean unsatisfactory performance.

In addition to counting the number of people, our vibration-based approach also provides fine-grained information on occupants’ footsteps (see Figure 16). Traditional people-counting methods are typically accomplished by scanning tickets or utilizing passive infrared (PIR) sensors to detect individuals passing through a doorway (Tsou et al., 2020), without providing any additional information regarding the status of those individuals. With vibration sensors, our method yields more detailed information about footstep dynamics than traditional people counting techniques, thereby offering a deeper insight into crowd behavior. For instance, this data can be employed to infer the walking speed of individuals, which is related to the risk of crowd-crushing accidents. A rapid walking pace, especially when coupled with a large number of individuals entering a space simultaneously, typically increases safety risks. Moreover, our approach can capture age-related variations in gait between older and younger individuals and facilitate the detection of disabilities, enabling staff to provide timely assistance to those in need. Furthermore, it assists in evaluating the health status of an individual’s gait, as outlined in existing research (Kessler et al., 2019; Fagert et al., 2021; Dong and Noh, 2024). Additionally, it offers insights

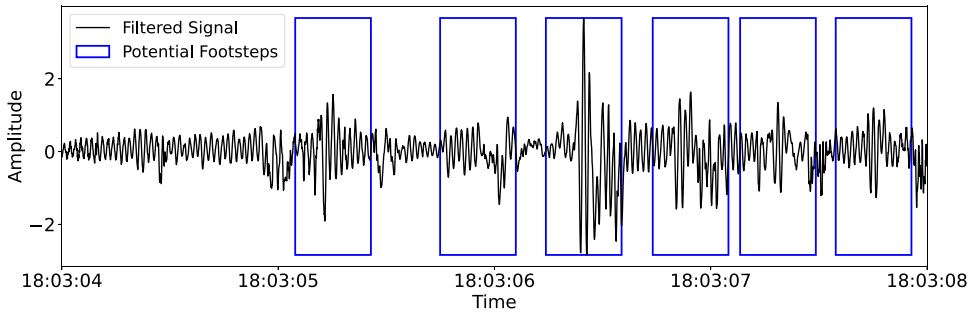


Figure 16. Potential footsteps captured during crowd traffic monitoring, demonstrating the capability of our approach to providing fine-grained footstep information such as crowd emotion in addition to counting people.

into the urgency of movements and can infer the emotional status of individuals (Wu et al., 2023), which helps improve the crowd experience.

4.3. Discussion on the effect of Variables in our approach

In this section, we discuss the variables that affect crowd behaviors in games, such as game types, sensor locations, promotional events, and crowd levels.

4.3.1. Effect of game types

The game types affect the distribution of crowd reaction and traffic, mainly through the popularity and intensity of the game, and the time when the game happens. To visualize the difference across games, we visualize the improvement of our method when compared to the data-only baseline models (see Figure 17).

At Stanford Maples Pavilion, women's basketball tends to attract a larger audience than men's and therefore leads to a higher level of crowd traffic and more intensive crowd reactions such as stomping. Moreover, most women's basketball games happen on weekend nights, which further increases the crowd traffic around the food stands and amplifies crowd reactions more than the games that happen in the afternoons. On the other hand, the football game at Michigan Stadium attracts $50 \times$ more people than the

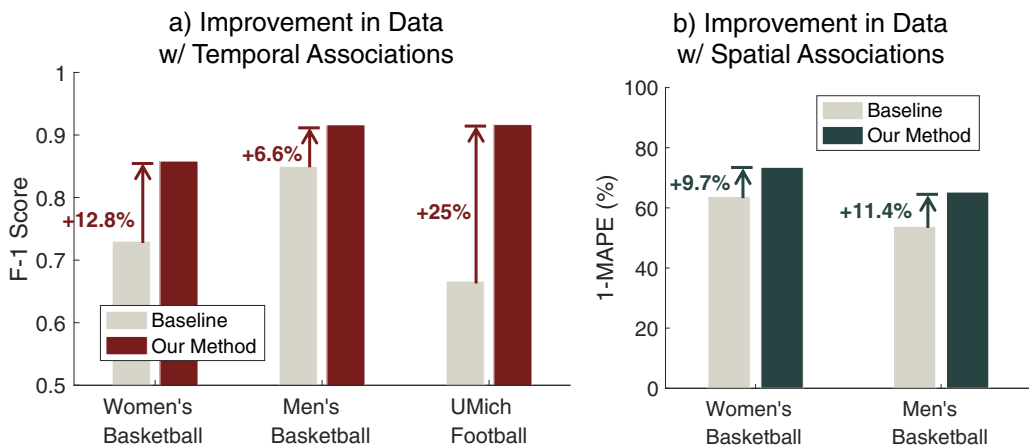


Figure 17. Our approach has significant improvement for both basketball games and football games, visualized for (a) data with temporal associations and (b) data with spatial associations.

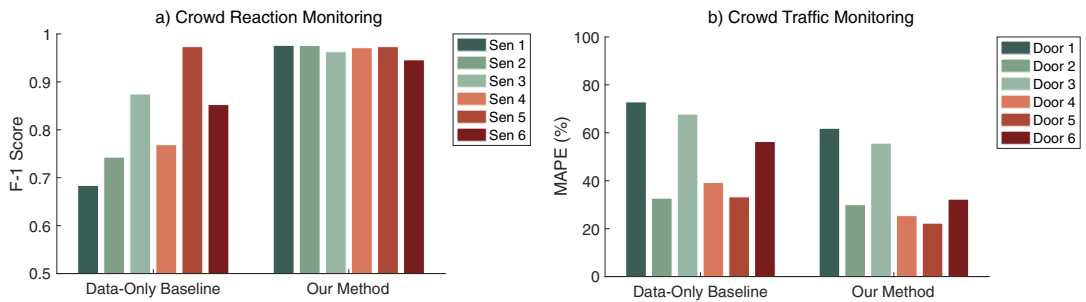


Figure 18. Effect of sensor locations in (a) crowd reaction monitoring and (b) crowd traffic monitoring. (a) Indicates that the temporal context (i.e., game progress) in our method corrects the between-sensor variations in the baseline. (b) Shows that the spatial context (i.e., facility layout) in our method does not have a significant impact on between-sensor variations.

Stanford site, which also leads to much noisier data which lowers the data-only performance. In this case, our approach significantly boosts the performance by introducing the game context, making it comparable to the smaller-scale basketball games.

To this end, we also observe the relatively consistent performance of our method in crowd monitoring performance across various game types. Figure 17 shows that while the baseline's performance varies significantly, our approach has only slight variations among basketball and football games. The difference in crowd reaction monitoring in the data-only baseline model is mainly due to the number of audiences, leading to noisier signals and more frequent loss of packets during data transmission, which will be further discussed in Section 4.3.4. For crowd traffic estimation, however, men's basketball games have a larger percentage error. This is because the size of the audience is smaller in men's games, resulting in a large MAPE as the MAE is divided by the overall smaller size of the audience entering each door during each 10-min interval. Our approach compensates for these issues through context association and modeling, leading to more balanced performance for all types of games.

4.3.2. Effect of sensor locations

The performance of crowd monitoring varies across sensor locations due to differences in crowd behavior heterogeneity and vibration data quality, influenced by factors such as the supported team, audience participation, surrounding noise levels, and floor/bleacher material properties. Figure 18 shows a performance comparison among various sensor locations for crowd reaction and traffic monitoring. The inconsistent performance in the data-only baseline indicates there are discrepancies in data quality across various sensor locations/doors. In Figure 18a, our context-aware approach mitigates this issue and produces better and more consistent performance in crowd reaction monitoring. In Figure 18b, while the error has been reduced by incorporating the spatial context, there is no significant improvement in the disparity among sensors. This means that the crowd traffic depends more on the quality of the sensor data than the spatial contexts and requires future work to further improve the data quality.

Another cause of the between-sensor disparity is the variability in crowd activities at various locations, which can also be shown in the crowd reaction type distribution around the stadium in Figure 19. A notable period of missed records aligns with the halftime interval across all six examined games due to the absence of the recorders. Despite these variations, there is a general consistency in activity types across different areas, attributed to the uniformity of audience engagement during game highlights. This is evidenced by the synchronization of movement and activity patterns; for example, increased movement is typically observed pre and postgame, as well as immediately before and after halftime, whereas heightened activity levels are predominantly during gameplay. However, activity frequency noticeably varies among areas, especially highlighted during the January 21, 2023 game, where areas monitored by sensors 1 and 6 experienced significantly more activity compared to those covered by sensors 2 and

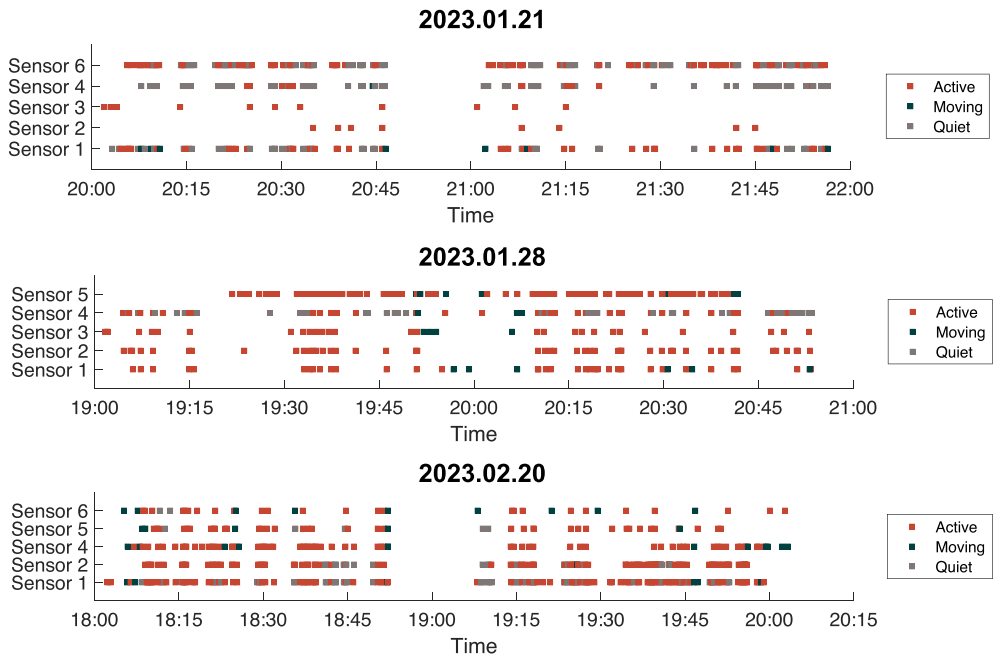


Figure 19. Comparison of crowd reaction type across multiple sensors over time. There is high variability in the crowd reactions among different sensor locations, with sensor 6 on January 21 and sensor 5 on January 28 demonstrating notably higher activity intensity.

3. Such variations are largely due to the proximity of sensors 1 and 6 to the student door, a hub for student gatherings. Observations indicate that students, drawn to these sectors, display greater enthusiasm and a higher rate of participation in activities throughout the game. Similarly, for the game on January 28, 2023, the audience in the area monitored by sensor 5 (the student area) exhibited a pattern of arriving late, engaging in more activities, and then leaving early before the game concluded.

4.3.3. Effect of promotional events

The promotional events such as free food, raffles, and entertainment sections organized by the stadium facilities for better crowd engagement affect the crowd traffic pattern. Figure 20 shows the effect of

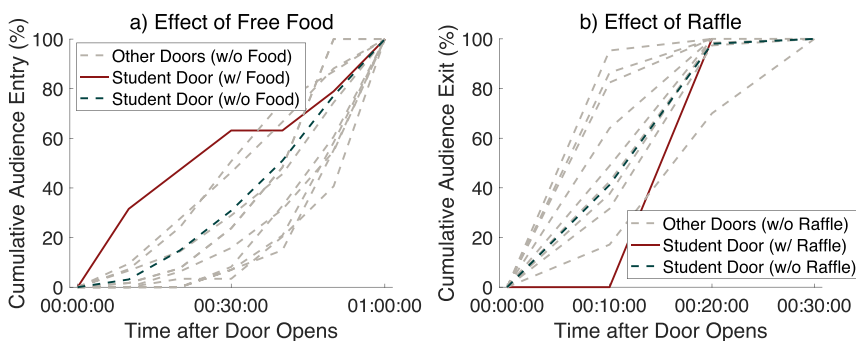


Figure 20. Effect of the promotional events on crowd traffic pattern at the doors, visualized for (a) free food served before the game, and (b) raffle prize after the game at the student door. We observe an initial rise in cumulative audience entry (%) and a delayed rise in cumulative audience exit (%) at the student door as compared to the student door and other doors without promotional events.

different promotional events on the crowd traffic pattern for entering and exiting the doors before and after the game respectively. Figure 20 (a) shows the initial rise in cumulative audience entry (%) at the student door when free food is served compared to the student door and other doors without free food events. During the free food event at the student door, people tend to come early into the game at the student door whereas, without the free food event, the crowd pattern at the student door is similar to other doors. Figure 20 (b) shows a delayed rise in cumulative audience exit (%) at the student door when raffle prizes are given compared to the student and other doors without raffle prizes events. During the raffle prize event at the student door, people tend to stay back after the game at the student door whereas, without the raffle prize event, the crowd pattern at the student door is similar to other doors with people leaving the doors as soon as the game ends. In the future, we plan to incorporate the impacts of these promotional events on crowd traffic estimation.

4.3.4. Effect of crowd levels

An unexpected challenge to wireless sensing of crowds was the effect of the crowd levels on data transmission. Take the Stanford Stadium as an example, four wireless routers were able to connect easily across the basketball court at the Stanford Maples Pavilion. However, with the addition of occupants, who both absorb radio waves and introduce electrical noise from devices on their person, the wireless backhaul connections between access points started to break down.

Figure 21 shows that the crowd density increases the vibration data loss during transmission from the vibration sensors to the routers. Figure 21(a) demonstrates the loss of data across outdoor sensors and Figure 21(b) demonstrates the loss of data across indoor sensors during the game. The highlighted regions in the figure correspond to the time when the door opens to the game begins (black), half-time (red), and the game ends to the door closes (blue). It can be observed that loss of vibration data increased when the doors opened and people started moving in across both indoor and outdoor sensors. Further, the loss of vibration data decreased drastically 20 min after the game ended as most of the people left the stadium, and crowd density was low.

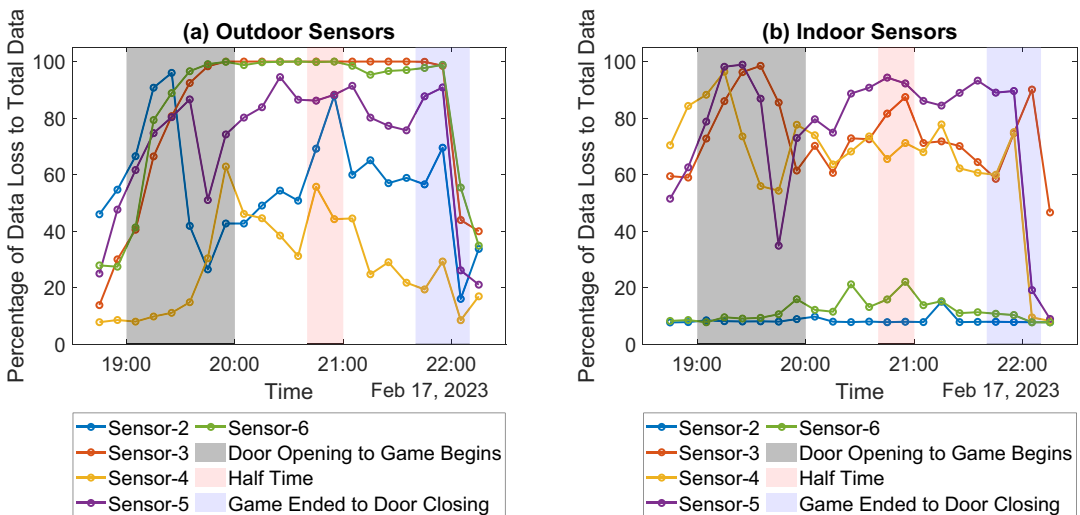


Figure 21. Percentage of data loss in indoor and outdoor sensors with respect to the crowd level changes over time. The highlighted regions represent the time when the crowd level changes significantly: (1) from door opening to game start (crowd level increases), (2) half-time (crowd level increases), and (3) from game end to door closing (crowd level decreases). This figure demonstrates the increase in crowd level during the game leads to more data loss in sensors.

Data loss can significantly impact the performance of our system. For instance, errors in crowd traffic estimation increase by 5–30% as crowd density rises, such as when more people enter the stadium. Conversely, sensors with minimal data loss (e.g., sensors 2 and 6 in [Figure 21\(b\)](#)) maintain consistent performance over time. It is important to note that system performance is influenced by multiple factors, including data quality and variations in crowd behavior. While no obvious correlation is observed between the system error and the data loss in our deployment, future work is recommended to examine their relationship to further improve the robustness of the system.

Fortunately, due to the self-organizing nature of 802.11s WiFi meshes ([Hiertz et al., 2010](#)), the number of access points is flexible, so we add additional routers to reduce the mesh hop distance, improving data transmission reliability. At the Michigan Stadium, we improved the sensor connection by adding more routers to the stadium, which has effectively reduced the amount of data loss for all sensors. In future deployments, our experience suggests that once connections are established in a given environment, the maximum hop distance should be halved to ensure robustness with large crowds.

5. Related work

To provide a background for this study, we review the existing literature on human/animal-induced structural vibration sensing and crowd monitoring through other sensing modalities.

5.1. Human- or animal-induced structural vibration sensing

The potential of using structural vibrations to infer behaviors of humans or animals has been explored in many previous studies. Our prior work has shown promise in using footstep-induced floor vibrations for occupancy detection ([Reuland et al., 2017](#); [Mirshekari et al., 2020](#)), person identification ([Pan et al., 2017](#); [Dong et al., 2023b](#)), gait analysis and physical health monitoring ([Alwan et al., 2003](#); [Kessler et al., 2019](#); [Dong and Noh, 2024](#); [Dong et al., 2024a, 2024b](#)). In addition to footsteps, vibrations induced by human activities can also be used for the prediction and characterization of activity types and patterns ([Alwan et al., 2006](#); [Pan et al., 2019](#); [Bonde et al., 2020](#)). Moreover, structural vibration sensing has been successful in animal health and activity monitoring ([Bonde et al., 2021](#); [Dong et al., 2023a](#)). These studies provide a knowledge base on how human/animal-induced structural vibrations can be used to infer their behaviors, which inspired the sensor deployment, feature extraction, and evaluation design of this study.

5.2. Crowd monitoring through other sensing modalities

Existing methods for crowd monitoring include manual monitoring, wearable devices, questionnaires, videos, audio recordings, WiFi and radio frequency signals, and so on. Manual monitoring is the most common approach for crowd monitoring. It is efficient and interpretable, but is labor-intensive, costly, and can be significantly delayed due to negligence in manual observations. Questionnaires are used for crowd monitoring. However, this method is also time-consuming and unable to gather timely information during the events ([Glass, 2005](#)). Centralized communication is introduced for immediate feedback ([Sgouros, 2000](#)). While it offers timely observation, the continuous messaging may intrude on attendees' experiences. One study utilized skin interfaces for crowd monitoring ([Stevens et al., 2009](#)), but they are required to be carried by each person and can be intrusive. Other works use cameras or microphones to catch crowd behaviors ([Maheshwari and Heda, 2016](#); [Bahmanyar et al., 2019](#)). However, these devices usually come with privacy concerns and thus may not be allowed in many public spaces. WiFi- and radio frequency-based devices are used to capture the body motion of individuals ([Yamin et al., 2018](#)). However, they have difficulty capturing the activity among a large group of people due to noise interference and between-person differences, producing less accurate results. Compared to the existing method, structural vibration sensing is non-intrusive, wide-ranged, less sensitive to loud sounds, and is perceived as more privacy-friendly, allowing continuous monitoring of crowd behavior in large, noisy indoor spaces.

5.3. Real-world impact of crowd monitoring

Crowd monitoring has a significant social impact to ensure public safety and can save lives. For example, 173 people lost their lives in the Bethnal Green disaster due to a stampede that occurred when a huge crowd was using stairs. Helbing et al. observed that bottleneck locations such as doors, narrow passways, and stairs are critical locations for crowd traffic monitoring where the probability of occurrence of a crowd disaster such as a stampede is higher (Helbing et al., 2000). Our crowd traffic estimation through floor vibration sensing can help stadium facilities make decisions regarding adequate resource allocation before, during, and after the game at each door to avoid crowd disaster situations. Further, crowd reaction monitoring for different sections of the stadium can prevent crowd disasters caused by critical situations such as panic, false alarms, and crowd clashes. These critical situations created in the section can cause a large number of people to move all at once as reported in panic incidents at Little War Memorial Stadium, 2018. Prolonged higher energy vibration signals created by the movement of a large number of people can help warn stadium facility authorities of the possible critical situation in the concerned section of the stadium.

6. Conclusions and future work

In this paper, we introduce a novel crowd-monitoring approach using crowd-induced floor vibrations. Compared with the existing approaches such as manual observation, sound-, video-, infrared-, and WiFi/RF-based sensing, our approach is cost-efficient, wide-ranged, and perceived as more privacy-friendly, which allows ubiquitous, fine-grained crowd behavior monitoring in public. We overcome the challenge of the high uncertainty in crowd-induced vibrations by developing a context-aware probabilistic model. This allows more accurate estimations of crowd reaction and traffic. We evaluate our approach through real-world deployments at Stanford Maples Pavilion and Michigan Stadium for six basketball games and two football games. Results show a 0.89 F-1 score and 9 MAE in crowd reaction and traffic monitoring, respectively. We also discussed the effect of game types, sensor locations, promotional events, and crowd levels.

For future work, we will improve our approach from three perspectives (1) sensing system design, (2) multi-sensor data modeling, and (3) downstream tasks with multimodal fusion. First, we will improve the sensing platform design based on the lessons learned from our experiments. This involves optimizing the number and placement of sensors and routers, as well as understanding the spatial resolution change with respect to the density of the sensors. Secondly, we will improve the context-aware modeling by incorporating uncertainties in the game progress records and facility layout due to delayed or missing information and variations in maps across different games. Overlapping activities in crowd settings, such as simultaneous walking and clapping will be considered and decoupled through the development of new algorithms. In addition, we plan to fuse multiple sensors by developing a graph neural network with sensors as nodes and the audience association as edges to provide more accurate crowd reaction inferences. Finally, we will target downstream applications such as predicting crowd-crushing risks and crowd emotion that relate to critical safety concerns. This information will provide live feedback to event managers to allocate the facility and manual resources optimally. Furthermore, we plan to integrate our vibration sensor networks with mobile/wearable devices and social media image/text data to enhance spatial and temporal resolution. This will lead to our ultimate goal of high-resolution, nonintrusive crowd monitoring with hybrid multimodal fusion.

Acknowledgments. The views and conclusions contained here are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either express or implied, of any University, Corporation, or the National Science Foundation. Special thanks to our volunteers who have shown remarkable generosity and dedication in recording ground truths for the games.

Data availability statement. Data will be provided upon reasonable request to the corresponding author.

Author contribution. Conceptualization: H.N.; P.Z.; H.L.; Y.D. Methodology: Y.D.; Y.W. Data curation: Y.D.; Y.W.; Y.C.; J.A.; J.C. Data visualisation: Y.D.; Y.W.; Y.C.; J.A. Writing original draft: Y.D.; Y.W.; Y.C.; J.A. All authors approved the final submitted draft.

Funding statement. This work was funded by the US National Science Foundation (under grant number NSF-CMMI-2026699), Stanford Blume Fellowship, Cisco Research, and Stanford CEE Graduate Fellowship.

Competing interest. The authors declare no competing interests exist.

Ethical standard. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

References

- Al-Shaery AM, Alshehri SS, Farooqi NS and Khoziun MO** (2020) In-depth survey to detect, monitor and manage crowd. *IEEE Access* 8:209008–209019.
- Alwan M, Dalal S, Kell S and Felder R** (2003) Derivation of basic human gait characteristics from floor vibrations. In *2003 Summer Bioengineering Conference*. IEEE, pp. 25–29.
- Alwan M, Rajendran PJ, Kell S, Mack D, Dalal S, Wolfe M and Felder R** (2006) A smart and passive floor-vibration based fall detector for elderly. In *2006 2nd International Conference on Information & Communication Technologies*, vol. 1. IEEE, pp. 1003–1007.
- Andersson M, Rydell J and Ahlberg J** (2009) Estimation of crowd behavior using sensor networks and sensor fusion. In *2009 12th International Conference on Information Fusion*. IEEE, pp. 396–403.
- Bahmanyar R, Vig E and Reinartz P** (2019) Mrcnet: crowd counting and density map estimation in aerial and ground imagery. Preprint, arXiv:1909.12743.
- Bonde A, Codling JR, Naruethep K, Dong Y, Siripaktanakorn W, Ariyadech S, Sangpetch A, Sangpetch O, Pan S, Noh HY and Zhang P** (2021) PigNet: failure-tolerant pig activity monitoring system using structural vibration. In *Proceedings of the 20th International Conference on Information Processing in Sensor Networks (co-located with CPS-IoT week 2021)*. Nashville, TN, USA: ACM, pp. 1–13.
- Bonde A, Pan S, Mirshekari M, Ruiz C, Noh HY and Zhang P** (2020) OAC: overlapping office activity classification through IoT-sensed structural vibration. In *2020 IEEE/ACM Fifth International Conference on Internet-of-Things Design and Implementation (IoTDI)*. IEEE, pp. 216–222.
- de Almeida MM and von Schreeb J** (2019) Human stampedes: an updated review of current literature. *Prehospital and Disaster Medicine* 34(1):82–88.
- Dong Y, Bonde A, Codling JR, Bannis A, Cao J, Macon A, Rohrer G, Miles J, Sharma S, Brown-Brandt T, et al.** (2023a) Pigsense: structural vibration-based activity and health monitoring system for pigs. *ACM Transactions on Sensor Networks* 20, 1–43.
- Dong Y, Fagert J, and Noh HY** (2023b) Characterizing the variability of footstep-induced structural vibrations for open-world person identification. *Mechanical Systems and Signal Processing* 204:110756.
- Dong Y, Iammarino M, Liu J, Codling J, Fagert J, Mirshekari M, Lowes L, Zhang P and Noh HY** (2024a) Ambient floor vibration sensing advances the accessibility of functional gait assessments for children with muscular dystrophies. *Scientific Reports* 14(1):10774.
- Dong Y, Kim SE, Schadh K, Huang P, Ding W, Rose J and Noh HY** (2024b) In-home gait abnormality detection through footstep-induced floor vibration sensing and person-invariant contrastive learning. *IEEE Journal of Biomedical and Health Informatics* 28(12).
- Dong Y and Noh HY** (2024) Ubiquitous gait analysis through footstep-induced floor vibrations. *Sensors* 24(8):2496.
- Dong Y, Wu Y, Codling JR, Aggarwal J, Huang P, Ding W, Latapie H, Zhang P and Noh HY** (2023c) Gamevibes: vibration-based crowd monitoring for sports games through audience-game-facility association modeling. In *Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. ACM, pp. 177–188.
- Fagert J, Mirshekari M, Pan S, Lowes L, Iammarino M, Zhang P and Noh HY** (2021) Structure-and sampling-adaptive gait balance symmetry estimation using footstep-induced structural floor vibrations. *Journal of Engineering Mechanics* 147(2): 04020151.
- Filingeri V, Eason K, Waterson P and Haslam R** (2017) Factors influencing experience in crowds—the participant perspective. *Applied Ergonomics* 59:431–441.
- Glass R** (2005) Observer response to contemporary dance. In *Thinking in Four Dimensions: Creativity and Cognition in Contemporary Dance*. Melbourne, Australia: Melbourne University Press Carlton, pp. 107–121.
- Haque S, Sadi MS, Rafi MEH, Islam MM and Hasan MK** (2020) Real-time crowd detection to prevent stampede. In *Proceedings of International Joint Conference on Computational Intelligence: IJCCI 2018*. Springer, pp. 665–678.
- Helbing, D., Farkas, I., and Vicsek, T.** (2000). Simulating dynamical features of escape panic. *Nature* 407(6803):487–490.
- Hiertz GR, Denteneer D, Max S, Taori R, Cardona J, Berlemann L and Walke B** (2010) IEEE 802.11s: the WLAN mesh standard. *IEEE Wireless Communications* 17(1):104–111.
- Jarvis N, Hata J, Wayne N, Raychoudhury V and Gani MO** (2019) Miamimapper: crowd analysis using active and passive indoor localization through wi-fi probe monitoring. In *Proceedings of the 15th ACM International Symposium on QoS and Security for Wireless and Mobile Networks*. ACM, pp. 1–10.
- Kantarci B and Mouftah HT** (2014) Trustworthy sensing for public safety in cloud-centric internet of things. *IEEE Internet of Things Journal* 1(4):360–368.

- Kessler E, Malladi VVS and Tarazaga PA** (2019) Vibration-based gait analysis via instrumented buildings. *International Journal of Distributed Sensor Networks* 15(10):1550147719881608.
- Kingshott BF** (2014) Crowd management: understanding attitudes and behaviors. *Journal of Applied Security Research* 9(3): 273–289.
- Kok VJ, Lim MK and Chan CS** (2016) Crowd behavior analysis: a review where physics meets biology. *Neurocomputing* 177: 342–362.
- Kumar A et al.** (2021) Crowd behavior monitoring and analysis in surveillance applications: a survey. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 12(7):2322–2336.
- Lamba S and Nain N** (2017) Crowd monitoring and classification: a survey. In *Advances in Computer and Computational Sciences: Proceedings of ICCCCS 2016*, vol. 1, Springer, pp. 21–31.
- Maheshwari S and Heda S** (2016) A review on crowd behavior analysis methods for video surveillance. In *Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies*. New York City, USA: ACM, pp. 1–5.
- Mirshekari M, Fagert J, Pan S, Zhang P and Noh HY** (2020) Step-level occupant detection across different structures through footstep-induced floor vibration using model transfer. *Journal of Engineering Mechanics* 146(3):04019137.
- Mirshekari M, Pan S, Fagert J, Schooler EM, Zhang P and Noh HY** (2018) Occupant localization using footstep-induced structural vibration. *Mechanical Systems and Signal Processing* 112:77–97.
- Molloy MS, Sherif Z, Natin S and McDonnell J** (2009) Management of mass gatherings. In Schultz CH and Koenig KL ed., *Koenig and Schultz's Disaster Medicine: Comprehensive Principles and Practices*. Cambridge, UK: Cambridge University Press, pp. 265–293.
- Pan S, Berges M, Rodakowski J, Zhang P and Noh HY** (2019) Fine-grained recognition of activities of daily living through structural vibration and electrical sensing. In *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. ACM, pp. 149–158.
- Pan S, Yu T, Mirshekari M, Fagert J, Bonde A, Mengshoel OJ, Noh HY and Zhang P** (2017) Footprintid: Indoor pedestrian identification through ambient structural vibration sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1(3):1–31.
- Reuland Y, Pai SG, Drira S and Smith IF** (2017) Vibration-based occupant detection using a multiple-model approach. In *Dynamics of Civil Structures, volume 2. Proceedings of the 35th IMAC, A Conference and Exposition on Structural Dynamics 2017*. Springer, pp. 49–56.
- Sgouros NM** (2000) Detection, analysis and rendering of audience reactions in distributed multimedia performance. In *Proceedings of the Eighth ACM International Conference on Multimedia*. ACM, pp. 195–200.
- Sharma A, McCloskey B, Hui DS, Rambia A, Zumla A, Traore T, Shafi S, El-Kafrawy SA, Azhar EI, Zumla A, et al.** (2023) Global mass gathering events and deaths due to crowd surge, stampedes, crush and physical injuries—lessons from the seoul halloween and other disasters. *Travel Medicine and Infectious Disease* 52.
- Stevens CJ, Schubert E, Morris RH, Frear M, Chen J, Healey S, Schoknecht C and Hansen S** (2009) Cognition and the temporal arts: investigating audience response to dance using pdas that record continuous data during live performance. *International Journal of Human-Computer Studies* 67(9):800–813.
- Tsou P-R, Wu C-E, Chen Y-R, Ho Y-T, Chang J-K and Tsai H-P** (2020) Counting people by using convolutional neural network and a PIR array. In *2020 21st IEEE International Conference on Mobile Data Management (MDM)*. IEEE, pp. 342–347.
- Wang J** (2021) A survey on crowd counting methods and datasets. In *Advances in Computer, Communication and Computational Sciences: Proceedings of IC4S 2019*. Springer, pp. 851–863.
- Wu Y, Dong Y, Vaid S, Harari GM and Noh HY** (2023) Emotion recognition using footstep-induced floor vibration signals. *Structural Health Monitoring* 2023.
- Yamin M, Basahel AM and Abi Sen AA** (2018) Managing crowds with wireless and mobile technologies. *Wireless Communications and Mobile Computing* 2018.
- Yogadhita G and Agustin W** (2023) Football stampede in Kanjuruhan stadium from the perspective of disaster preparedness on mass casualty incident: a case study of mass gathering event. *Prehospital and Disaster Medicine* 38(S1):s78–s79.
- Zeitz KM, Tan HM, Grief M, Couns P and Zeitz CJ** (2009) Crowd behavior at mass gatherings: a literature review. *Prehospital and Disaster Medicine* 24(1):32–38.
- Zhang G, Lu D and Liu H** (2018) Strategies to utilize the positive emotional contagion optimally in crowd evacuation. *IEEE Transactions on Affective Computing* 11(4):708–721.