GameVibes: Vibration-based Crowd Monitoring for Sports Games through Audience-Game-Facility Association Modeling

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

Yuyan Wu
Stanford University
Stanford, USA

Jesse R Codling
University of Michigan
Ann Arbor, USA

Jatin Aggarwal
Stanford University
Stanford, USA

Peide Huang
Carnegie Mellon University
Pittsburgh, USA

Wenhao Ding
Carnegie Mellon University
Pittsburgh, USA

Hugo Latapie
Cisco Systems, Inc.
San Jose, USA

Pei Zhang
University of Michigan
Ann Arbor, USA

Hae Young Noh
Stanford University
Stanford, USA

Figure 1: Crowd reactions during a NCAA Pac-12 Basketball Game at Stanford Maples Pavilion.

ABSTRACT

Crowd monitoring involves tracking and analyzing the behavior of large groups of people in large-scale public spaces, such as sports games. In sports stadiums, understanding audience reactions to the games and their distribution around the public facilities is important for ensuring public safety and security, enhancing the game experience, and improving crowd management. Recent crowd-crushing incidents (e.g., Kanjuruhan Stadium disaster, Seoul Halloween Stampede) have caused 100+ deaths in a single event, 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 *GameVibes*, a novel method for crowd behavior monitoring using crowd-induced floor vibrations to infer audience reactions to the game (e.g., clapping, stomping, dancing) and crowd traffic (i.e., the number of people entering each door). The main benefits of *GameVibes* are that it allows continuous, fine-grained crowd monitoring in a cost-effective and non-intrusive way and is perceived as more privacy-friendly. Unlike monitoring an individual person, crowd monitoring involves understanding the overall behavior of a large population (typically more than 1,000), leading to high uncertainty in the vibration data. To overcome the challenge, we first establish the game and facility association to inform the context of crowd behaviors, including 1) *game associations* (temporal context) between the crowd reaction and the game progress and 2) *facility associations* (spatial context) between the crowd traffic and facility layouts. Then, we formulate the crowd monitoring problem by converting the conceptual graph of the

BuildSys '23, November 15–16, 2023, Istanbul, Turkey

ACM ISBN 979-8-4007-0230-3/23/11…$15.00
https://doi.org/10.1145/3600100.3623750

audience-game-facility association into probabilistic game/facility association models. Through these models, *GameVibes* first learns the latent representations of the game progress and facility layout through neural network encoders, and then integrates heterogeneous game/facility information and vibration data to estimate crowd behaviors. This mitigates the estimation error due to the uncertainty in vibration data. To evaluate our approach, we conduct 6 real-world deployments for NCAA Pac-12 games at Stanford Maples Pavilion. Our results show that *GameVibes* achieves a 0.9 F-1 score in crowd reaction monitoring and 9.3 mean absolute error in crowd traffic estimation, which correspond to 10% and 12.2% error reduction, respectively, compared to the baseline methods without context-specific information.

KEYWORDS

crowd behavior, floor vibration, context, association, sports game

ACM Reference Format:
Yiwen Dong, Yuyan Wu, Jesse R Codling, Jatin Aggarwal, Peide Huang, Wenhao Ding, Hugo Latapie, Pei Zhang, and Hae Young Noh. 2023. GameVibes: Vibration-based Crowd Monitoring for Sports Games through Audience-Game-Facility Association Modeling. In *The 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '23), November 15–16, 2023, Istanbul, Turkey.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3600100.3623750

1 INTRODUCTION

Crowd monitoring is the process of tracking and analyzing the behavior of large groups of people in large-scale public spaces, such as sports games and shopping malls [24]. Especially, crowd monitoring in sports stadiums is a critical component in ensuring public safety and security [37], enhancing the game experience for the audience [13], and optimizing resource allocation at the stadiums [21, 28]. Over the years, the 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 [33, 36]. Nevertheless, studies have found that such incidents can be prevented by proper crowd monitoring and timely crowd control [8]. By analyzing the reactions and traffic of the crowd, we can detect and prevent potential threats, such as riots, stampedes, violence, or terrorist attacks [18]. Moreover, we can also gain insights into the social and psychological aspects of crowd behavior, such as emotions and motivations, to understand the fundamental cause behind such behaviors [38].

While many existing approaches are developed to monitor crowd behaviors, few meet the practical needs due to their limitations in cost, privacy protection, and accuracy [1, 4, 15, 19, 23]. For example, manual monitoring is the most common approach [22], which is labor-intensive, costly, and can be significantly delayed due to negligence in manual observations. Automatic devices such as video and audio data suffer from privacy issues due to the appearance and voice recordings of the public [5, 25]. On the other hand, WiFi- and radio frequency-based devices are used to capture the body motion of individuals [17, 35], but they have difficulty in capturing the activity among a large group of people due to noise interference and between-person differences, producing inaccurate results.

In this paper, we introduce *GameVibes*, a novel system for crowd monitoring using vibration sensors mounted underneath or on the floor surfaces. *GameVibes* captures crowd-induced vibrations to infer crowd reactions to the game (e.g., clapping, stomping, dancing) and crowd traffic (i.e., the number of people entering each door). The primary intuition behind *GameVibes* is that various types of crowd behaviors and levels of crowd traffic induce distinct vibration patterns of the floor, allowing us to characterize and distinguish crowd behaviors. The main benefits of using floor vibration to infer human behaviors are that it is cost-efficient, allows continuous, fine-grained crowd behavior 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 [26, 31], identification [10, 30], activity recognition [3, 29], localization [27], and health monitoring [11, 12, 20].

However, crowd monitoring is not a trivial problem because it involves a large group of people (typically more than 1,000), which causes high uncertainty in crowd-induced vibration data. Unlike monitoring an individual person, crowd monitoring involves understanding the overall behavior of a large population, which involves 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 possible number of pedestrians. As a result, estimating crowd behaviors may lead to much larger errors than monitoring an individual person.

To overcome the challenges, we leverage the audience-game-facility associations, which bridge the context of the game/facility with crowd reactions and traffic. Specifically, *GameVibes* establishes 1) *game associations* (temporal context) between the crowd reaction and the game progress, such as clapping after the home team's goals, stomping to disturb the opponents' free throws, and 2) *facility associations* (spatial context) between the crowd traffic and facility layouts, such as crowd accumulates around the entry doors near the food stands. With the established associations, we formulate the crowd monitoring problem by converting the conceptual graph of the audience-game-facility association into probabilistic game/facility association models. Through these models, *GameVibes* first learns the latent representations of the game progress and facility layout through neural network encoders, and then merge the heterogeneous game/facility information and vibration data to estimate the crowd behaviors. With the audience-game-facility association modeling, *GameVibes* mitigates the estimation error due to the uncertainty in the vibration data, leading to more accurate and interpretable crowd monitoring.

The key contributions of this paper are:

- We introduce *GameVibes*, the first floor-vibration-based crowd monitoring system that continuously monitors crowd reactions during sports games and estimates crowd traffic over various entry locations.
- We characterize the game- and facility-dependent floor vibrations induced by the crowd behaviors to develop game and facility association models, which provide temporal and spatial contexts to the vibration data to allow more accurate and interpretable crowd monitoring.
- We evaluate the *GameVibes* system through 6 real-world deployments for NCAA sports games at Stanford Maples

Pavilion, which validates its effectiveness and robustness under various scenarios.

For the rest of this paper, we first characterize crowd-induced vibrations with game and facility contexts to formulate the audience-game-facility association model (Section 2), then introduce the components of the *GameVibes* system (Section 3). Next, we present the 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 CHARACTERIZING CROWD-INDUCED VIBRATION WITH GAME AND FACILITY CONTEXTS

In this section, we characterize the relationship among crowd behaviors, vibration data, and game/facility contexts to develop an audience-game-facility association model for crowd reaction and traffic estimation. Specifically, we first discuss how crowd reaction is affected by game progress and reflected in floor vibration of the bleachers (Section 2.1), and then analyze how crowd traffic is influenced by facility layout and captured by floor vibration (Section 2.2). Then, we establish game and facility associations to develop probabilistic graphical models that merge heterogeneous information (game/facility information and vibration data) to make context-aware estimations of crowd behaviors.

2.1 Crowd Reaction Characterization through Floor Vibration and Game Progress

To understand how crowd reaction is affected by game progress and reflected in floor vibration, we first characterize the vibration signals with respect to crowd reaction to validate the vibration-based approach for crowd reaction monitoring. Then, we analyze the relationship between the game progress and the crowd reactions in order to leverage this relationship as a temporal context to the vibration data in Section 2.3.

2.1.1 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 2. 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.

2.1.2 Relationship between Crowd Reaction and Game Progress. With the relationship between floor vibration and crowd reactions, we further analyze how audience reaction changes as the game progresses. Figure 3 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

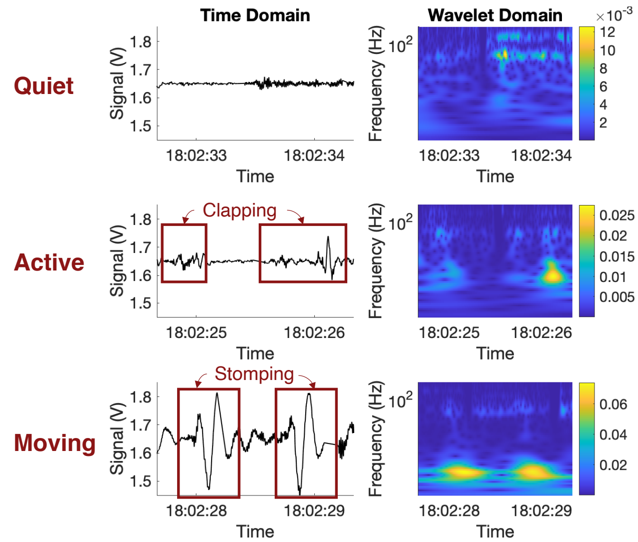

Figure 2: Characterization of vibration signals induced by crowd reactions, including quiet, active, and moving (from top to bottom). Both time- and wavelet-domain plots show clear distinctions among various crowd reaction types.

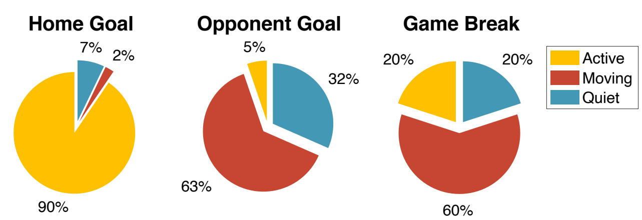

Figure 3: Crowd reaction distribution varies across various events at a sample game. Active (clapping) dominates the home goal event, while moving (stomping) and quiet dominate the opponent goal event. The reactions are more evenly distributed during the game break.

clapping (i.e., active) after the home goal, while remaining quiet 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 kids' mini-game, T-shirt toss, and dance cams.

2.2 Crowd Traffic Characterization through Floor Vibration and Facility Layout

To validate the vibration-based approach for crowd traffic estimation, we first characterize the relationship between floor vibration and crowd traffic. Then, We analyze how facility layout affects crowd traffic in order to leverage this influence as a spatial context to the vibration data in Section 2.3.

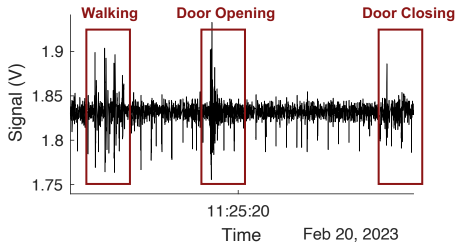

Figure 4: Floor vibrations captured by a sample sensor near an entry door. The impulsive peaks induced by walking and opening or closing the door are detected by a peak-picking algorithm, which correlates with the level of crowd traffic.

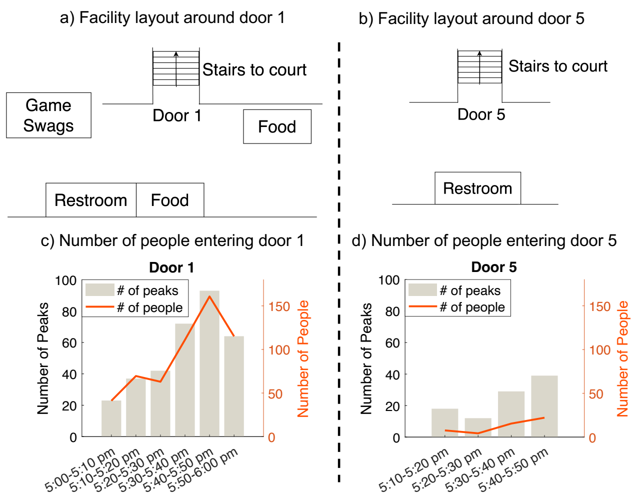

Figure 5: The difference in facility layout around a) door 1 and b) door 5 leads to distinct crowd traffic as shown in c) and d). The crowd traffic is correlated with the number of peaks detected in the vibration signals.

2.2.1 Relationship between Crowd Traffic and Floor Vibration. The number of audience entering through each door can be inferred from the floor vibration signals captured by sensors deployed at the floor's surface beside each door. The physical intuition is that the movements of the audience passing through the door, including walking and opening or closing the door, generate vibration in the floor structures. These vibrations are then detected and recorded by sensors mounted on the floor surface (refer to Figure 4). We notice that there is a positive correlation between the number of peaks and the number of people entering the doors (See Figure 5). This indicates the frequency of audience movements (e.g., walking, door opening/closing) correlates with the number of people entering through each door during a time interval. For example, more footsteps and door-opening events will be observed when the crowd traffic is of higher volume. Therefore, detecting impulsive peaks in the recorded vibration signals induced by audience movements allows crowd traffic estimation at each door.

2.2.2 Relationship between Crowd Traffic and Facility Layout. The location and distribution of essential facilities such as food stations, restrooms, and game swags play a crucial role in determining crowd traffic. For example, Figure 5 shows the facility layout around two sample doors at our evaluation site, 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 5c) 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 2.3.

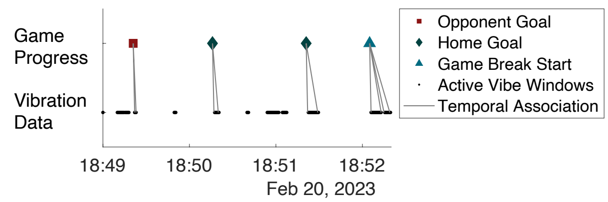

Figure 6: Game association examples between crowd-induced vibrations and the game progress. Each game event such as the opponent goal, home team goal, and game break start is matched with active vibration windows over time.

2.3 Formulation of the Audience-Game-Facility Association Models

To incorporate the influence of game progress and facility layout on crowd behaviors, we establish game/facility associations to provide temporal and spatial contexts to the vibration data. Specifically, we establish 1) the *game association* between crowd reaction and the game progress to provide a temporal context, and 2) the *facility association* between the crowd traffic and facility layout to provide a spatial context. With these contexts, we formulate the game/facility association models to allow more accurate and interpretable estimation of crowd behaviors.

2.3.1 Establishing Game Associations. The *game 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 *game 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-second windows for discrete association. Figure 6 shows a snapshot of the *game 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. 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.

2.3.2 Establishing Facility Association. The *facility association* refers to the relationship between the vibration data recorded at entry doors and the facilities around those doors based on spatial distance. For example, if a sensor is placed at an entry door with a food stand nearby, then the vibration data reflects the crowd traffic around the food stand, which has a distinct traffic pattern when compared to a door without any facilities.

We establish the *facility associations* through location proximity between the sensor and each facility type. The facilities we consider include restrooms, game swag stations, food stands, and a student center. Facilities within a certain walking distance near the doors are associated with the corresponding sensor, in which the distance threshold is chosen as the overall maximum distance to reach any facility starting from the closest door. The distance is important because it reflects the strength a *facility association* - the shorter the distance is, the more influence a facility may have on the crowd traffic. These associations are encoded as a spatial proximity matrix to enhance the accuracy of our crowd traffic estimation. The details of how we leverage the facility layout for crowd traffic estimation are described in Section 3.4.2.

2.3.3 Developing Audience-Game-Facility Association Model for Crowd Monitoring. With the game (temporal context) and facility (spatial context) associations established in the previous subsections, we formulate the crowd monitoring problem by developing a game association and a facility association model that formalize the relationship among vibration data, game progress, and facility layout. As summarized in Figure 7, the conceptual graphs (left) describing audience-game-facility relationships are converted into the corresponding probabilistic graphical models (right), allowing probabilistic analysis of the crowd behavior through vibration data. Specifically, we formulate the game and facility association model for crowd reaction monitoring (upper) and crowd traffic estimation (lower), respectively.

Crowd Reaction Monitoring Formulation: The upper left part of Figure 7 shows the conceptual graph between crowd reaction and game progress. According to the discussion in Section 2.1, the game progress affects the crowd reaction, with which the vibration data can be associated through the sequence of occurrence in time. To this end, we formulate a probabilistic game association model among crowd reaction (Y), game progress (G), and vibration data (X) based on the conceptual graph, where dependencies and game associations are maintained. Assuming that the game record is an accurate and timely reflection of the game progress, G is regarded as a deterministic variable in our model. With this formulation, the objective of crowd reaction monitoring is to estimate $P(Y|X, G)$.

Crowd Traffic Estimation Formulation: Similarly, the discussion in Section 2.2 shows that the layout of the facility at each door affects crowd traffic, so the vibration data collected at that door can be spatially associated with the surrounding facilities (see the lower part of Figure 7). Based on this conceptual relationship, we formulate a probabilistic facility association model among crowd traffic (Y), facility layout (F), and vibration data (X), where the dependencies and facility associations are maintained. Given that the proximity of each type of facility around each door is known, F is also regarded as a deterministic variable in our model. To this end, the objective of crowd traffic estimation is to compute $P(Y|X, F)$.

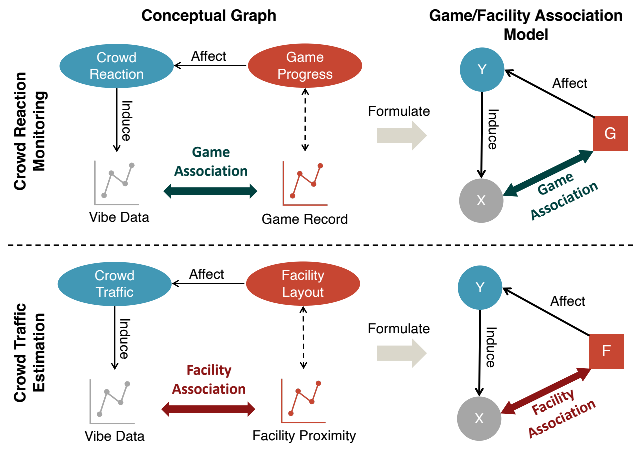

Figure 7: Conceptual graph (left) and the corresponding probabilistic game/facility association model (right). The upper half describes the relationships among the crowd reaction (Y), game progress (G), and vibration data (X) through the *game association*; the lower half summarizes the relationships among the crowd traffic (Y), facility layout (F), and vibration data (X) through the *facility association*. G and F are in squared boxes, representing deterministic variables.

3 CROWD BEHAVIOR MONITORING THROUGH AUDIENCE-GAME-FACILITY ASSOCIATION MODELING

In this section, we first provide an overview of our *GameVibes* system and then present each module in the system for crowd monitoring throughout sports games.

3.1 Overview of *GameVibes*

Our *GameVibes* system consists of three modules: 1) Sensing and Data Pre-progressing, 2) Game-informed Crowd Reaction Monitoring, and 3) Facility-informed Crowd Traffic Estimation. The input of the *GameVibes* system is the crowd-induced vibration, the game record, and the facility layout in the stadium.

In the first module, we collect the vibration data through vibration sensors mounted underneath the floor (for bleachers) or attached to the floor surface (at the entry doors). The vibration signals are transmitted wirelessly to a centralized server and stored in a hard drive. We also pre-process the raw signals through sliding windows and interpolation algorithms to produce discretized signal segments for analysis.

In the crowd reaction monitoring module, we integrate the game record with the processed vibration data through the game association model introduced in Section 2.3. The module estimates crowd

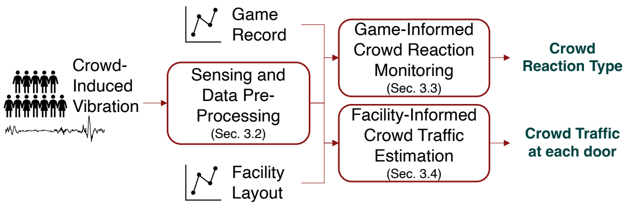

Figure 8: Overview of the *GameVibes* system in 3 modules. *GameVibes* integrates crowd-induced vibration, game record, and facility layout to estimate crowd reaction and traffic.

Figure 9: Sensor network of *GameVibes*. Sensor data are transmitted wirelessly through routers via Wi-Fi connections in the stadium.

reaction types, including audience status (e.g., quiet, active, moving) and the specific type of reaction under each status (e.g., clapping, stomping), which will be discussed in Section 3.3.

Similarly, for crowd traffic estimation, we integrate the facility layout with the processed vibration data based on the facility association model in Section 2.3. This module estimates the crowd traffic in terms of the number of people entering each door, which will be discussed in Section 3.4.

3.2 Sensing and Data Pre-processing

GameVibes's geophone-based vibration sensing platform is developed based on the design that has been successful in previous animal and human welfare applications [6, 9]. This platform consists of robust, independent geophone sensing nodes that communicate over a private WiFi network. 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.

The setting of the large-scale sports stadium requires additional adaptations to the previous sensor network design. Compared to existing vibration-based sensing platforms, *GameVibes* represents a significant increase in scale, both in the deployment area and the number of occupants. Because of this scale-up, many of the assumptions made in our previous sensing network [6, 9] no longer applied. For example, the mains-powered nodes running for months at a time were changed to battery-powered geophone sensors which can operate for at most 24 hours. This tradeoff was deemed acceptable for the stadium environment where individual games are relatively short, on the order of several hours, and the network can be re-deployed for each event being monitored.

Operating in a larger area with stricter deployment requirements informed changes to the network topology as well. Previously, a simple star network was in use [6, 9], as the default for WiFi-based systems. Since this was now insufficient, a multi-hop mesh network, as shown in Figure 9, was used. 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 for the public location.

An unexpected challenge to wireless sensing of crowds was the impact of the crowds themselves on data transmission. During testing, four wireless routers were able to connect easily across the basketball court. 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. Fortunately, due to the self-organizing nature of 802.11s Wi-Fi meshes [16], the number of access points is flexible, so we add additional routers to reduce the mesh hop distance, improving data transmission reliability. 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.

After collecting the vibration data, we process the signals to prepare for window-based analysis and mitigate the missing data issue for subsequent data modeling. First, we segment the signals into 1-second windows with a time step of 0.5 seconds 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.

3.3 Game-informed Crowd Reaction Monitoring

In this section, we introduce the module for crowd reaction monitoring, which integrates the game record and vibration data to monitor crowd reactions. Each step of this module will be discussed in the next few subsections based on Figure 10.

3.3.1 Crowd Reaction Detection and Feature Extraction. We first detect vibration windows that capture audience motions using the processed vibration data (noted as Vibe Data). Crowd reaction detection is performed by comparing the signal window to the noise signal. To capture the noise characteristics, we select signal windows for approximately 2 minutes during periods of inactivity (the time when the stadium is empty). These windows serve as the noise signal for reference. Assuming a normal distribution for the noise, we calculate the mean and standard deviation values for these noise signals for each sensor. The signals are then subtracted from the mean value of the noise signal to maintain its zero average. To accommodate the high noise variance caused by the loud music during the game (approximately 20 times the standard deviation of the noise according to our observations), we choose 20 times of the standard deviation of the noise signal as the threshold for crowd reaction detection. Windows with amplitudes surpassing

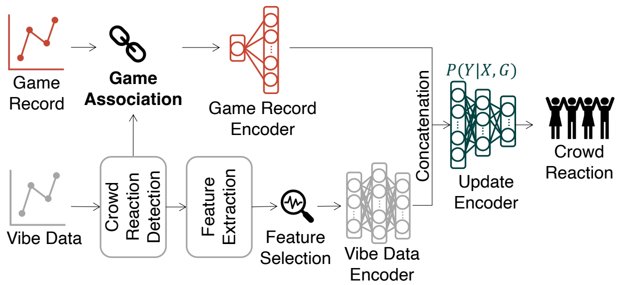

Figure 10: Module for crowd reaction monitoring in *GameVibes*. Game associations are first established, and then encoded by neural networks to extract latent representations for the game and vibe data. Finally, the representations of the game record and vibration data are integrated through an update encoder to estimate crowd reactions.

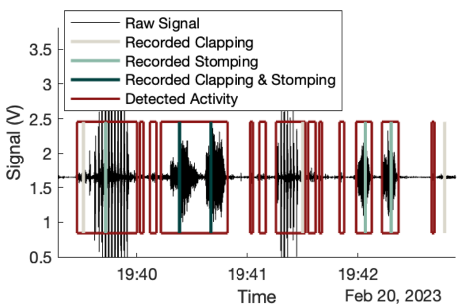

Figure 11: Activity detection results on sample vibration data. The algorithm successfully detects 83% of crowd reactions within the 5-second error range of the recorded reactions.

this threshold are detected as windows with audience motion. Subsequently, we smooth these identified activities across adjacent 5 windows through moving median to refine the detection and enhance the accuracy of our analysis. As shown in Figure 11, this algorithm successfully detects 83% of crowd reactions within the 5-second error range of the manually labeled reactions.

After the crowd reaction detection, we extract features from the vibration signals, summarized as follows:

- *Time-domain Feature*: signal energy of each 0.1-s segment.
- *Frequency-domain Feature*: cumulative 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-second grid block.

These features provide a more comprehensive description of the crowd reactions than a single domain, covering multiple aspects of the vibration signals, such as its variations in time, frequency, and dependencies between time and frequency. These features are then selected by a random forest model to rank their importance to the crowd reaction. 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 effective information in vibration data.

3.3.2 Modeling of Game Record and Vibration Data. We first establish the *game associations* between the game record and the crowd-induced vibration data, and then develop two different encoders, for modeling of game record and vibration data, respectively.

Game Association Establishing: The *game associations* are established between the game record and the active windows of vibration data based on Section 2.3.1. With the associations, the game progress (G) and the vibration data (X) are linked through time, enabling game-informed crowd reaction monitoring.

Game and Vibration Data Modeling: To model the game record and vibration data, we design two neural networks with different characteristics to encode the features. First, we leverage a game record encoder to learn the latent variables representing the multifaceted influence of the game progress on audience-induced vibrations (e.g., intensity, duration) by expanding a one-dimensional score change or game break indicator to a multi-dimensional 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 1-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 3-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.

3.3.3 Audience-Game Integration for Crowd Reaction Monitoring. After modeling the game record and vibration data, we concatenate their learned embeddings and integrate this information through an update encoder to estimate the conditional distribution $P(Y|X,G)$ as introduced in Section 2.3.3. The encoder is a 3-layer funnel-shaped neural network that gradually transforms the concatenated embeddings to approximate the distribution of $P(Y|X,G)$. The resultant conditional probability is represented as a vector with the same length as the number of audience reaction types.

3.4 Facility-informed Crowd Traffic Estimation

In this section, we introduce the module for crowd traffic estimation by integrating the facility layout and vibration data to estimate the crowd traffic (i.e., headcounts at each door). The next few subsections will present each step in Figure 12.

3.4.1 Crowd Traffic Feature Extraction and Data Augmentation. We conduct feature extraction based on the floor vibration signals to capture information related to the level of crowd traffic. As illustrated in Figure 4 of Section 2.2.1, the peaks of the floor vibration signal represent the movements of the audience passing through the door including walking, opening, or closing the door. Therefore, we detect these peaks and extract features from these peaks to estimate the crowd traffic. First, we identify peaks by setting a minimum peak distance of 1 second 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

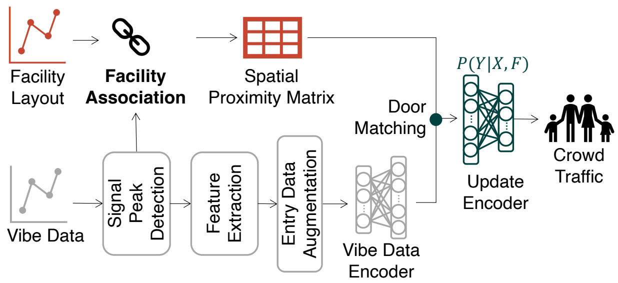

Figure 12: Module for crowd traffic estimation in *GameVibes*. Facility associations are first established and then modeled by a spatial proximity matrix. Then, we integrate the spatial proximity matrix and the vibration data through door matching and an update encoder to estimate the crowd traffic.

determined through a preliminary analysis of the training data. The chosen threshold corresponds to the highest correlation score observed during this analysis. Then, we extracted features based on the detected peaks to relate floor vibration to capture information related to the level of crowd traffic, summarized as follows:

- *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.

To overcome the limited size of our dataset, we augment our sample by merging two 10-minute 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 aggregating the number of people entering within these two windows.

3.4.2 Modeling of Facility Layout and Vibration Data. We establish the *facility associations* between the facility layout and sensors at various doors and then leverage a spatial proximity matrix and an encoder neural network to model facility layout and vibration data, respectively.

Facility Association Establishing: The *facility associations* are established between the facility layout and the sensor at each door based on the stadium map provided by the venue operator, as discussed in Section 2.3.2. With these associations, the facility layout (F) and vibration data (X) are linked through their locations.

Facility Proximity and Vibration Data Modeling: To model the facility proximity and vibration data, we develop 1) a spatial proximity matrix and 2) a vibration data encoder neural network,

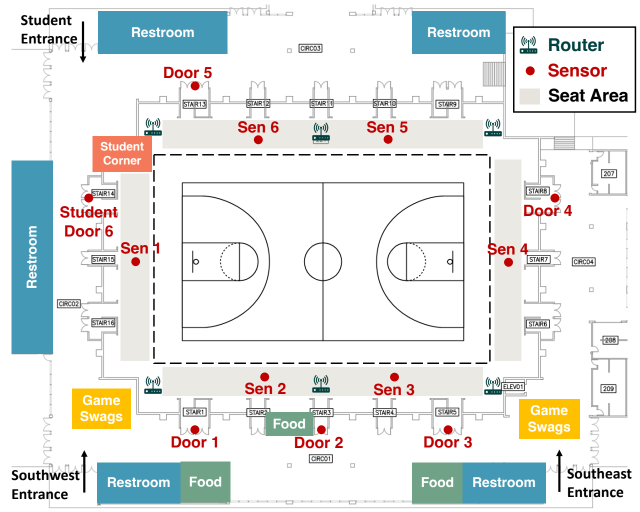

Figure 13: Experiment setup at Stanford Maples Pavilion with the sensor layout (marked as red dots), router layout (marked as green devices), facility locations at the concourse area (described as squares of different colors), and 16 entry doors connecting the concourse and the game court.

respectively. 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 meters in our case) and the actual distance between the facility and the door. A higher weight means a shorter distance to the facility, indicating a stronger association. These facilities include food stands, game swag stations, restrooms, and game courts, as described in Figure 13. For each facility type, a higher weight represents closer proximity to that door. On the other hand, the vibe data encoder is a 2-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.

3.4.3 Audience-Facility Integration for Crowd Traffic Estimation. 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 facility-aware updating. The update encoder is a 2-layer neural network that approximates the distribution of $P(Y|X, F)$. The output is the number of people entering each door.

4 REAL-WORLD EVALUATION AT STANFORD MAPLES PAVILION

To evaluate *GameVibes*, we conduct 6 real-world deployments for NCAA pac-12 women's and men's basketball games at Stanford Maples Pavilion, producing more than 280 hours of vibration data from 12 sensors. In this section, we first introduce the deployment setup, and then show the results for crowd monitoring. Furthermore, we discuss the variables that affect crowd behaviors and results, including game types, sensor locations, and promotional events.

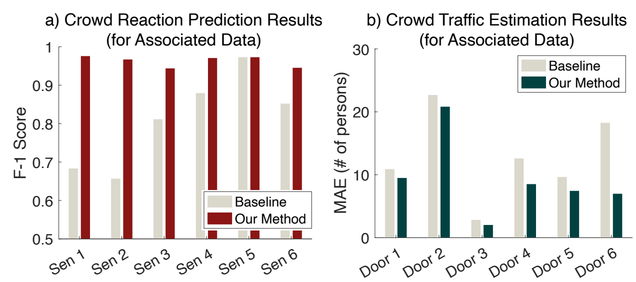

Figure 14: Our *GameVibes* system outperforms the baseline at all sensor locations for a) crowd reaction monitoring (higher F-1 score), and b) crowd traffic estimation (lower MAE).

4.1 Deployment Setup

The deployment setup is shown in Figure 13, 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 3.2. The sampling frequency is set to 500 Hz to maximize the temporal resolution while ensuring data transmission efficiency.

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

4.2 Overall Performance of *GameVibes*

Overall, *GameVibes* has a 0.9 F-1 score in crowd reaction monitoring and 9.3 mean absolute error (MAE) in estimating headcounts for crowd traffic estimation among various doors. Compared to the baseline methods without audience-game-facility association, *GameVibes* has averages of 10% and 12.2% improvements, respectively. The results are summarized in Figure 14. The performance increase is mainly because *GameVibes* incorporates the temporal and spatial contexts through the game/facility associations with the vibration data. As the game progresses drive the crowd reactions and facility layout direct the crowd traffic, these contexts provide reliable prior information for crowd monitoring.

4.2.1 Crowd Reaction Monitoring Performance. *GameVibes* has a 0.9 and 0.83 F-1 score in audience status and reaction classification, respectively. To show the effectiveness of the game association, we also compare the overall performance with the performance on windows that have game associations, which has an average of 10% improvement in the F-1 score. The improvement indicates that the game context corrects multiple misclassified samples due to the highly uncertain data. This is because the latent representations learned from the game record can indirectly reflect the cause of crowd reaction variations, such as the active crowd size and the activity intensity.

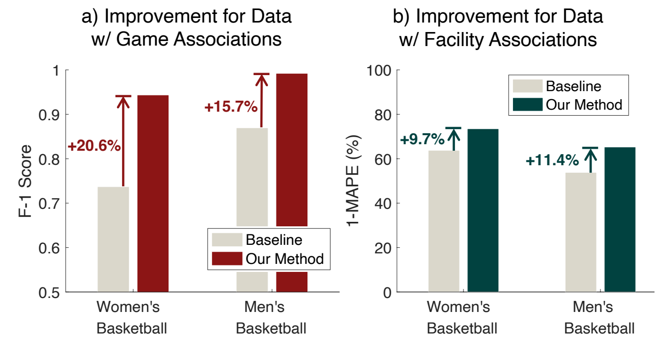

Figure 15: *GameVibes* has up to 20.6% and 11.4% improvement for women's and men's basketball games, visualized for a) data with game associations, and b) data with facility associations.

4.2.2 Crowd Traffic Estimation Performance. *GameVibes* has an average of 9.3 MAE for crowd traffic estimation, which has an average of 12.2% improvement among all doors when compared to the baseline method without facility association. To understand the relative error for crowd traffic estimation, we also compute the mean absolute percentage error (MAPE) by averaging MAEs for all the 10-min periods for all games, which is 30.6% on average for all doors. 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 minutes of entry.

4.3 Discussion of Variables in *GameVibes*

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

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. Based on our observation, 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.

We also observe variations in *GameVibes*'s crowd monitoring performance across various game types. Figure 15 shows that *GameVibes* has slightly different performance for women's and men's basketball games. The lower performance for crowd reaction monitoring is due to the 2× more audience in women's games, leading to noisier signals and more frequent loss of packets during data transmission. For crowd traffic estimation, however, men's 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

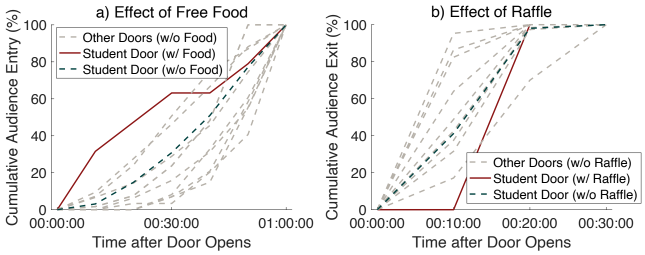

Figure 16: Effect of the promotional events on crowd traffic pattern, visualized for a) free Food before the game, and b) raffle after the game at the student door. We observe an earlier increase in cumulative audience entry (%) and a later increase in cumulative audience exit (%) at the student door as compared to other doors without promotional events.

entering each door during each 10-minute interval. Our *GameVibes* compensates for these issues through audience-game-facility association modeling, leading to more balanced performance for both types of games.

4.3.2 Effect of Sensor Locations. The sensor locations affect the performance of crowd monitoring mainly through the vibration data quality, which is determined by various factors, such as the noise in the surroundings, the floor/bleacher material property, and the level of interference during data transmission. The inconsistent performance in baseline indicates there are discrepancies in data quality across various sensor locations/doors (see Figure 14). In contrast, our method with audience-game-facility association mitigates this issue and produces better and more consistent performance, especially for crowd reaction monitoring.

4.3.3 Effect of Promotional Events. The promotional events include free food, raffles, and entertainment sessions to engage the audience before and after the game. These events mainly affect the crowd traffic. For example, the student door sometimes has free food and raffles before/after the game. People tend to arrive early to the game on the day when free food is served as observed by the high initial increase in cumulative audience entry (%) at the student door in Figure 16a). When there is no free food, cumulative audience entry (%) follows a similar pattern to other doors. Further, people tend to stay back to collect raffles at the end of the game as observed by a late increase in cumulative audience exit (%) in Figure 16b). When there is no raffle event, cumulative audience exit (%) follows a similar pattern to other doors with most people leaving the door as soon as the game ends. We plan to study the impact of these promotional events and incorporate their effect on crowd traffic estimation in future work.

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/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 the footstep-induced floor vibrations for occupancy detection [26, 31], identification [10, 30], gait health monitoring [2, 11, 12, 20]. In addition to footsteps, vibrations induced by human activities can also be used for the prediction and characterization of activity types and patterns [3, 7, 29]. Moreover, structural vibration sensing has been successful in animal health and activity monitoring [6, 9]. 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 [14]. Centralized communication is introduced for immediate feedback [32]. While it offers timely observation, the continuous messaging may intrude on attendees' experiences. One study utilized skin interfaces for crowd monitoring [34], but they are required to be carried by each person and can be intrusive. Other works use cameras or microphones to catch crowd behaviors [5, 25]. 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 [35]. 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.

6 CONCLUSIONS AND FUTURE WORK

In this paper, we introduce *GameVibes*, a novel system for vibration-based crowd monitoring. *GameVibes* is cost-efficient, wide-ranged, and perceived as more privacy-friendly, which allows ubiquitous, fine-grained crowd behavior monitoring in the public. We overcome the challenge of the high uncertainty in crowd-induced vibrations by modeling audience-game-facility associations. This allows context-aware and more accurate estimations of crowd reaction and traffic. We evaluate *GameVibes* through real-world deployments for 6 NCAA sports games at Stanford Maples Pavilion and achieve 0.9 F-1 score and 9.3 MAE in crowd reaction monitoring and crowd traffic estimation, respectively.

For future work, we will first improve the audience-game-facility association model by incorporating uncertainties in the game progress records and facility layout due to delayed or missing information and variations in maps across different games. We will also explore larger and more complex scenarios (e.g., football games) and consider the distribution of the audience who support each team.

In addition, we will target downstream applications such as recognizing the crowd emotion, and detecting and predicting events concerning public safety.

ACKNOWLEDGMENTS

This work was funded by the U.S. National Science Foundation (under grant number NSF-CMMI-2026699), Cisco Systems, Inc., and Stanford CEE Graduate Fellowship. 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. They are: Akash Doshi, Andrew Jensen, Akhil Kode, Helen Zhang, Kai Kirk, Jingxiao Liu, Mo Wu, Sanat Mehta, and Yun Ni.

REFERENCES

[1] Ali M Al-Shaery, Shroug S Alshehri, Norah S Farooqi, and Mohamed O Khozium. 2020. In-depth survey to detect, monitor and manage crowd. *IEEE Access* 8 (2020), 209008–209019.

[2] Majd Alwan, Siddharth Dalal, Steve Kell, Robin Felder, et al. 2003. Derivation of basic human gait characteristics from floor vibrations. In *2003 Summer Bioengineering Conference, June*. 25–29.

[3] Majd Alwan, Prabhu Jude Rajendran, Steve Kell, David Mack, Siddharth Dalal, Matt Wolfe, and Robin Felder. 2006. A smart and passive floor-vibration based fall detector for elderly. In *2006 2nd International Conference on Information & Communication Technologies*, Vol. 1. IEEE, 1003–1007.

[4] Maria Andersson, Joakim Rydell, and Jorgen Ahlberg. 2009. Estimation of crowd behavior using sensor networks and sensor fusion. In *2009 12th International Conference on Information Fusion*. IEEE, 396–403.

[5] Reza Bahmanyar, Elenora Vig, and Peter Reinartz. 2019. MRCNet: Crowd counting and density map estimation in aerial and ground imagery. *arXiv preprint arXiv:1909.12743* (2019).

[6] Amelie Bonde, Jesse R. Codling, Kanittha Naruethep, Yiwen Dong, Wachirawich Siripaktanakon, Sripong Ariyadech, Akkarit Sangpetch, Orathai Sangpetch, Shijia Pan, Hae Young Noh, and Pei Zhang. 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)*. ACM, Nashville TN USA, 1–13. https://doi.org/10.1145/3412382.3458902

[7] Amelie Bonde, Shijia Pan, Mostafa Mirshekari, Carlos Ruiz, Hae Young Noh, and Pei Zhang. 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, 216–222.

[8] Maria Moitinho de Almeida and Johan von Schreeb. 2019. Human stampedes: An updated review of current literature. *Prehospital and disaster medicine* 34, 1 (2019), 82–88.

[9] Yiwen Dong, Amelie Bonde, Jesse R Codling, Adeola Bannis, Jinpu Cao, Asya Macon, Gary Rohrer, Jeremy Miles, Sudhendu Sharma, Tami Brown-Brandl, et al. 2023. PigSense: Structural Vibration-based Activity and Health Monitoring System for Pigs. *ACM Transactions on Sensor Networks* (2023).

[10] Yiwen Dong, Jonathon Fagert, and Hae Young Noh. 2023. Characterizing the variability of footstep-induced structural vibrations for open-world person identification. *Mechanical Systems and Signal Processing* 204 (2023), 110756.

[11] Yiwen Dong, Joanna Jiaqi Zou, Jingxiao Liu, Jonathon Fagert, Mostafa Mirshekari, Linda Lowes, Megan Iammarino, Pei Zhang, and Hae Young Noh. 2020. MD-Vibe: physics-informed analysis of patient-induced structural vibration data for monitoring gait health in individuals with muscular dystrophy. In *Adjunct proceedings of the 2020 ACM international joint conference on pervasive and ubiquitous computing and proceedings of the 2020 ACM international symposium on wearable computers*. 525–531.

[12] Jonathon Fagert, Mostafa Mirshekari, Shijia Pan, Linda Lowes, Megan Iammarino, Pei Zhang, and Hae Young Noh. 2021. Structure-and sampling-adaptive gait balance symmetry estimation using footstep-induced structural floor vibrations. *Journal of Engineering Mechanics* 147, 2 (2021), 04020151.

[13] Victoria Filingeri, Ken Eason, Patrick Waterson, and Roger Haslam. 2017. Factors influencing experience in crowds–the participant perspective. *Applied ergonomics* 59 (2017), 431–441.

[14] Renee Glass. 2005. Observer response to contemporary dance. *Thinking in four dimensions: Creativity and cognition in contemporary dance* (2005), 107–121.

[15] Sabrina Haque, Muhammad Sheikh Sadi, Md Erfanul Haque Rafi, Md Milon Islam, and Md Kamrul Hasan. 2020. Real-time crowd detection to prevent stampede. In *Proceedings of International Joint Conference on Computational Intelligence: IJCCI 2018*. Springer, 665–678.

[16] Guido R. Hiertz, Dee Denteneer, Sebastian Max, Rakesh Taori, Javier Cardona, Lars Berlemann, and Bernhard Walke. 2010. IEEE 802.11s: The WLAN Mesh Standard. *IEEE Wireless Communications* 17, 1 (Feb. 2010), 104–111. https://doi.org/10.1109/MWC.2010.5416357

[17] Nicholas Jarvis, John Hata, Nicholas Wayne, Vaskar Raychoudhury, and Md Osman Gani. 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*. 1–10.

[18] Wang Jingying. 2021. A survey on crowd counting methods and datasets. In *Advances in Computer, Communication and Computational Sciences: Proceedings of IC4S 2019*. Springer, 851–863.

[19] Burak Kantarci and Hussein T Mouftah. 2014. Trustworthy sensing for public safety in cloud-centric internet of things. *IEEE Internet of Things Journal* 1, 4 (2014), 360–368.

[20] Ellis Kessler, Vijaya VN Sriram Malladi, and Pablo A Tarazaga. 2019. Vibration-based gait analysis via instrumented buildings. *International Journal of Distributed Sensor Networks* 15, 10 (2019), 1550147719881608.

[21] Brian F Kingshott. 2014. Crowd management: Understanding attitudes and behaviors. *Journal of Applied Security Research* 9, 3 (2014), 273–289.

[22] Ven Jyn Kok, Mei Kuan Lim, and Chee Seng Chan. 2016. Crowd behavior analysis: A review where physics meets biology. *Neurocomputing* 177 (2016), 342–362.

[23] Ajay Kumar et al. 2021. Crowd behavior monitoring and analysis in surveillance applications: a survey. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 12, 7 (2021), 2322–2336.

[24] Sonu Lamba and Neeta Nain. 2017. Crowd monitoring and classification: a survey. In *Advances in Computer and Computational Sciences: Proceedings of ICCCCS 2016, Volume 1*. Springer, 21–31.

[25] Saurabh Maheshwari and Surbhi Heda. 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*. 1–5.

[26] Mostafa Mirshekari, Jonathon Fagert, Shijia Pan, Pei Zhang, and Hae Young Noh. 2020. Step-level occupant detection across different structures through footstep-induced floor vibration using model transfer. *Journal of Engineering Mechanics* 146, 3 (2020), 04019137.

[27] Mostafa Mirshekari, Shijia Pan, Jonathon Fagert, Eve M. Schooler, Pei Zhang, and Hae Young Noh. 2018. Occupant localization using footstep-induced structural vibration. *Mechanical Systems and Signal Processing* 112 (2018), 77–97. https://doi.org/10.1016/j.ymssp.2018.04.026

[28] Michael S Molloy, Zane Sherif, Stan Natin, and John McDonnell. 2009. Management of mass gatherings. *Koenig and Schultz's Disaster Medicine: Comprehensive Principles and Practices; Schultz, CH, Koenig, KL, Eds* (2009), 265–293.

[29] Shijia Pan, Mario Berges, Juleen Rodakowski, Pei Zhang, and Hae Young Noh. 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*. 149–158.

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

[31] Yves Reuland, Sai GS Pai, Slah Drira, and Ian FC Smith. 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, 49–56.

[32] Nikitas M Sgouros. 2000. Detection, analysis and rendering of audience reactions in distributed multimedia performance. In *Proceedings of the eighth ACM international conference on Multimedia*. 195–200.

[33] Avinash Sharma, Brian McCloskey, David S Hui, Aayushi Rambia, Adam Zumla, Tieble Traore, Shuja Shafi, Sherif A El-Kafrawy, Esam I Azhar, Alimuddin Zumla, 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 (2023).

[34] Catherine J Stevens, Emery Schubert, Rua Haszard Morris, Matt Frear, Johnson Chen, Sue Healey, Colin Schoknecht, and Stephen Hansen. 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 (2009), 800–813.

[35] Mohammad Yamin, Abdullah M Basahel, and Adnan A Abi Sen. 2018. Managing crowds with wireless and mobile technologies. *Wireless Communications and Mobile Computing* 2018 (2018).

[36] Gde Yogadhita and Widiana Agustin. 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 (2023), s78–s79.
[37] Kathryn M Zeitz, Heather M Tan, M Grief, PC Couns, and Christopher J Zeitz. 2009. Crowd behavior at mass gatherings: a literature review. *Prehospital and disaster medicine* 24, 1 (2009), 32–38.
[38] Guijuan Zhang, Dianjie Lu, and Hong Liu. 2018. Strategies to utilize the positive emotional contagion optimally in crowd evacuation. *IEEE Transactions on Affective Computing* 11, 4 (2018), 708–721.