Urban Space Insights Extraction using Acoustic Histogram Information

Nipun Wijerathne, Billy Pik Lik Lau[*], Benny Kai Kiat Ng, Chau Yuen

Singapore University of Technology and Design (SUTD), 8 Somapah Rd, Singapore 487372

Abstract

Urban data mining can be identified as a highly potential area that can enhance the smart city services towards better sustainable development especially in the urban residential activity tracking. While existing human activity tracking systems have demonstrated the capability to unveil the hidden aspects of citizens' behavior, they often come with a high implementation cost and require a large communication bandwidth. In this paper, we study the implementation of low-cost analogue sound sensors to detect outdoor activities and estimate the raining period in an urban residential area. The analogue sound sensors are transmitted to the cloud every 5 minutes in histogram format, which consists of sound data sampled every 100ms (10Hz). We then use wavelet transformation (WT) and principal component analysis (PCA) to generate a more robust and consistent feature set from the histogram. After that, we performed unsupervised clustering and attempt to understand the individual characteristics of each cluster to identify outdoor residential activities. In addition, on-site validation has been conducted to show the effectiveness of our approach.

Keywords: Smart City, Activity Detection, Sound Histogram, Sound Sensor, Clustering.

[*]Corresponding author

Email addresses: `hnipun@gmail.com` (Nipun Wijerathne), `billy_lau@mymail.sutd.edu.sg` (Billy Pik Lik Lau), `benny_ng@sutd.edu.sg` (Benny Kai Kiat Ng), `yuenchau@sutd.edu.sg` (Chau Yuen)

1. Introduction

Smart cities allow integrated systems to monitor the human and environmental behaviors in real-time have been looked upon as a potential solution to fulfill the requirements of modern urbanization. These statistics related to behaviors and knowledge are then used to make more effective decisions to further improve the urban design and living standards of people. With the recent enhancements of wireless networks, large-scale systems for various sectors can be realized easily, which resides an enormous amount of spatial-temporal data that can be harvested. These data contain many interesting aspects of day-to-day utilization of an urban space or life of citizens[1, 2, 3, 4, 5, 6]. With the advancement of machine learning and data mining technique, different platforms have been developed to cater different usages of monitoring in a smart city. Various aspects have to be taken into consideration when deploying a smart city project, such as deployment and maintenance cost, system and data scale, sensor coverage and resolution, user privacy, data storage, etc. Take all these factors into design and implementation consideration, urban monitoring in a smart city is extremely challenging. The increasing population density has driven more efficient real estate planning and at the same time needs to balance out aspects of liveable space in a dense area. Hence, the urban designer needs to understand the residential behavior of the existing urban area. Common technique such as outdoor activity detection is often used and it related to occupancy detection (can be indoor also). Such activity detection can be identified using sensors such as thermal imaging [7], camera [8, 9], motion[6, 10], mobile devices[11, 12, 13], and sound sensor [14, 15, 16]. A lot of the aforementioned techniques require high computation cost or communication bandwidth, which may not be suitable for continuous smart city monitoring. Not to mention, privacy intrusion is also another concern among researchers [17, 18].

In this paper, we focus on discovering human outdoor activity detection techniques and patterns in an urban point of interest (PoI) using low-cost analogue sound sensors. We want to show that using a low-cost analogue sensor is able to capture human outdoor activity while maintaining the sustainability of a sensor node and reducing communication overhead. However, the challenge in using the sound sensor is that

the outdoor environment tends to have a larger variance in ambient noise and hence it is harder to distinguish outdoor activity as well as background noise. Also, another huddle in collecting sound data is the sampling rate of the sensor, which may consume more power as well as constant communication between the gateway to transmit data. Thus, our motivation is to design a system model that process the histogram data from sound sensors to extract insights such as human activity and raining period of an urban PoI through compressed information such as sound histograms.

We leverage the idea of edge computation to achieve self-sustainability of the sensor node (using energy harvesting) at the same time to minimize the communication bandwidth between sensor nodes and gateway. The analogue sound sensor capture sound intensity collected from the surrounding environment, which can be composed of people chatting, children playing, pass-by vehicles, rain, etc. The sound data collected are complied at the sensor node into a sound histogram and uploaded to the cloud database every 5-minute for off-line processing. We first preprocess the sound histogram data using wavelet transformation (WT) and principal component analysis (PCA) to extract a more robust feature sets. WT has been extensively studied in area of speech recognition [19], outliers detection [20], and image processing [21]. Based on the literature, it seems like a natural selection to use WT for developing a richer feature space for sound histogram by allowing changes in the temporal aspect. Next, we use PCA to convert the transformed sound histogram into a set of uncorrelated values, which can be further utilized for insights extraction. PCA is studied in [22, 23, 24] and shows effective in generating uncorrelated values while reducing data dimension for easier visualization. Here, we divide the insights extraction process into two parts, which are raining period estimation and outdoor activity detection. The raining period is an example of a common event impact multiple sensors when compared to outdoor activity, where the event only can be detected by a single sensor. In addition, the raining period estimation is important in outdoor activity detection as raining day affects outdoor public space utilization. Similar principle can be extended to detect large area events such as explosion or party of a large group of participants, where multiple sensors are impacted simultaneously. To detect outdoor activity, we utilize the clustering method to group sound samples with high similarity and attempt to understand

3

background sound according to the time of the day and day of the week. Through the identified background sound, the different period has been defined and chi-square statistic is used for distinguishing the background sound and activity. The key challenge here is to classify the human outdoor activity and background sound accurately with a simple and effective method. Ground-truth data is also collected for validating the system model and show its effectiveness in a real-world scenario.

The contributions of our work in this paper are three-fold:

- We introduce the concept of compressing the sound samples of an urban environment as a histogram leveraging edge-computation techniques.

- We propose a system with preprocessing (WT and PCA) as well as clustering techniques to obtain useful feature set for identifying background noise and human outdoor activity. We show that the proposed system model is capable to achieve accuracy of 85.8% distinguishing background sound and human outdoor activity.

- Using multiple sensor nodes, we can estimate events that cover large areas, where we use the raining period estimation as an example, to show that using our proposed method we can get a better estimation than using a resistive water droplet sensor.

The remaining of this paper is organized as follows: Section 2 presents the system for processing the sound data as well as the details of the sound histogram data. Next in Section 3, we discuss the details used for preprocessing the sound histogram and methods for clustering processed sound samples. Section 4 describes the raining period estimation and human outdoor activity detection method, followed by Section 5, which validates our outdoor detection method against ground-truth. Lastly, we conclude our work in Section 6.

2. System Model

The system model consists of three different stages, which are (1) data collection phase, (2) data preprocessing and features generation phase, as well as (3) insight ex-

4

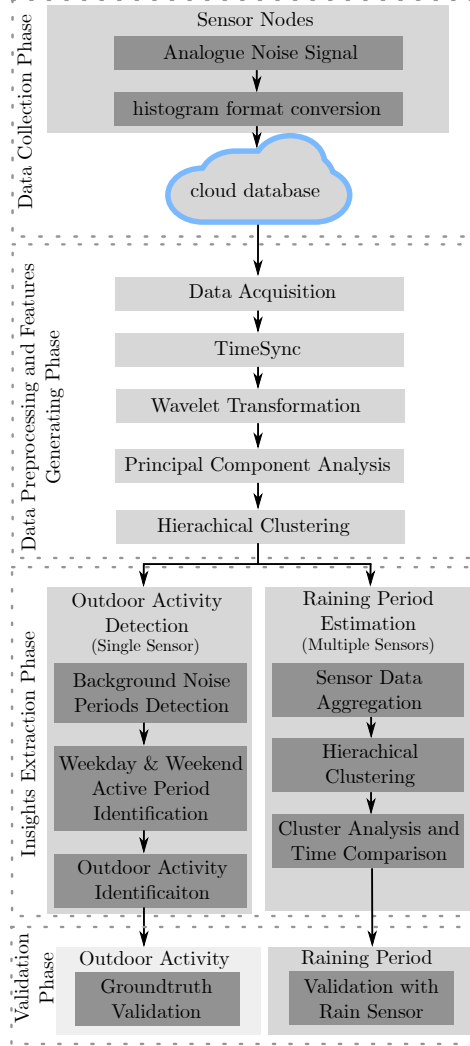traction phase. The overall system model is shown in Fig. 10.

Figure 1: Sound Data Processing System Model

The first stage involves data collection using sensor nodes in an urban residential area, which is first introduced by [2, 6]. The following describes the methodology of generating sound histograms from sound samples: we use an analogue sound sensor (dfrobot Analog Sound Sensor SKU: DFR0034) and it is capable of detecting the ambient sound intensity using LM386 amplifier and electret microphone. The output

of the sensor (noise intensity level) is an analog value, and we use a 10-bit ADC to quantize it into 1024 levels. It covers the range of 0 to 1023 with a higher value that indicates higher noise intensity. The sound sensor data is then sampled in the rate of every 100ms (10Hz sampling frequency). For every 5 minutes, we would have 3000 sound samples collected and this will be fitted into histogram $h(m)$.

$$h(m) = [s_1, ..., s_m] \tag{1}$$

where $\mathbf{h}$ is denote as the vector notation. s_m is the mth bin value and $m \in \{1, ..., 5\}$. These 3000 sound samples are divided into 5 bins of a histogram based on the value ranges. The rationality of choosing 5 bins of the histogram is that the majority of idling time (without any human activity), noise intensity has a lower range of value. Therefore, any higher value denotes chances of activity presence regardless of value number (i.e. value of 50 vs 333), which varies on noise source distance. Also, through the number of bin values, we can observe the duration of human activity. The following Table 1 show the range of quantifying the 0 to 1023 noise value:

Table 1: Sound Sensor Intensity Ranges

Range 1	Range 2	Range 3	Range 4	Range 5
$0 \leq i \leq 6$	$6 < i \leq 10$	$10 < i \leq 20$	$20 < i \leq 50$	$i > 50$

To further elaborate the consolidation process of sound samples, an example as follows is presented: Using 1 second sampling time as a baseline, we have a 10 sound samples collected, such as $[11, 2, 7, 33, 55, 80, 28, 7, 9, 13]$. Based on the collected samples, we consolidate them into a histogram of $[1, 3, 2, 2, 2]$. The rationality of choosing this method is to help to cope with the communication bandwidth requirements and reduces the communication overhead. Whenever the histogram has been consolidated, it will be uploaded to the cloud database via a wireless sensor network gateway every 5 minutes.

The second processing stage involves preprocessing techniques previously introduced in [2, 6] such as data acquisition and TimeSync. After performing cleaning up on the data, cleaned up histogram data are further processed to generate more features

for insights extraction. In this paper, we focus on two kinds of insights extraction from the sound sensor's histograms data format, which are (1) human outdoor activity detection and (2) raining period estimation. However, the former will be given more focus as it is the main purpose of this paper - to detect human outdoor activity in an urban area using analogue sound sensor. The latter will only be briefly discussed as we discover that it is possible to estimate the raining period using the histogram sound data aggregated from multiple sensors. Further details of both insights extraction methods will be elaborated in the upcoming session.

3. Sound Histograms Preprocessing and Clustering Phase

In this section, we discuss the in-depth techniques used to process the sound histogram. First, we explain the WT approach applied to transform the sound histograms into feature spaces. Subsequently, we utilize the PCA method to generate richer feature sets from WT applied sound histograms.

3.1. Wavelet Transformation

WT based feature extraction method is used to develop the feature space for sound histograms. Wavelet coefficients for a sound histogram are computed using a series of dilation and translation of the mother wavelet [25, 26]. We can represent the mth bin of a sound histogram as follows:

$$h(m) = \sum_{j} \sum_{k} d_{j,k} \psi_{j,k}(m),\qquad(2)$$

where $\psi_{j,k}(m)$ is the scaled and dilated mother wavelet function, the $d_{j,k}$ is the wavelet coefficient that represents how much translated and dilated mother wavelet describes the given sound histogram $h(m)$, and $j, k \in \mathbb{Z}$. from there, we can formulate the following equation:

$$d_{j,k} = \sum_{m} h(m)\psi_{j,k}(m) = < h, \psi_{j,k} >,\qquad(3)$$

since wavelets are orthogonal to each other.

Wavelets are simply mathematical functions that exhibit only localized oscillations, which decay quickly. This allows us to study local undulations of the sound histograms sound sample. For example, if the given sound histogram $h(m)$ has a discontinuity, then it will only influence the $\psi_{j,k}(m)$ that is in the vicinity. Only those coefficients $\{d_{j,k}\}_{j,k\in\mathbb{Z}}$ whose associated wavelet $\psi_{j,k}(m)$ overlaps the discontinuity will be influenced. Moreover, it should be noted that the bases of other common transformations are affected by the discontinuity, regardless of where it is located.

The localized behavior of the WT is important in generating the feature set for the histograms sound data. Thus, we use the *Haar* basis function as the mother wavelet due to its discontinuous nature. It can be used to analyze histograms effectively and efficiently at various resolutions and used to get the approximation coefficients and detail coefficients at various levels as shown in [25]. Besides, it is also served as a low-pass filter and a high-pass filter simultaneously. The *Haar* basis function can de denoted as follows:

$$\psi_{j,k}(x) = \begin{cases} 1 & \text{if } x \in [0, 1/2) \\ -1 & \text{if } x \in [1/2, 1) \\ 0 & \text{otherwise} \end{cases} . \tag{4}$$

In our system model, 5 minutes sound histogram sound samples will be represented in *Haar* basis function, and those wavelet coefficients $\{d_{j,k}\}_{j,k\in\mathbb{Z}}$ are used as the feature vector. Subsequently, discrete *Haar* wavelet coefficients can be calculated as

$$\mathbf{w} = \mathbf{B}^{-1}\mathbf{h}, \tag{5}$$

where $\mathbf{B}$ is the wavelet basis matrix and $\mathbf{w}$ is the wavelet coefficient vector for the sound histogram. Being restricted to dyadic sequences, traditional *Haar* wavelet construction is not sufficient. The reason due to the number of sound histogram bins is non-dyadic. Therefore, the intuition behind this method is to use a non-dyadic *Haar* construction, which is more accurate than the zero paddings.

Let us denote the elements at level i, interval j by $T_{i,j}$ and $|T_{i,j}| = n_{i,j}$. Here, the elements can be consists of the following: $T_{0,1} = \{s_1, s_2, s_3, s_4, s_5\}$, $T_{1,1} = \{s_1, s_2, s_3\}$, $T_{1,2} = \{s_4, s_5\}$, $T_{2,1} = \{s_1, s_2\}$, and $T_{2,2} = \{s_3\}$. The length of intervals being $n_{0,1} = 5, n_{1,1} = 3, n_{1,2} = 2, n_{2,1} = 2, n_{2,2} = 1$. We can represent the

elements of **B** that corresponds to detailed coefficients as follows:

$$B_{l,m} = \begin{cases} (\sqrt{n}\dfrac{1}{2}\sqrt{\dfrac{n_a + n_b}{n_a n_b}})\dfrac{2n_b}{n_a n_b} & \text{if } s_m \in T_{i+1,k} \\[2em] -(\sqrt{n}\dfrac{1}{2}\sqrt{\dfrac{n_a + n_b}{n_a n_b}})\dfrac{2n_a}{n_a n_b} & \text{if } s_m \in T_{i+1,k} \\[3em] 0 & \text{otherwise} \end{cases} \tag{6}$$

where the elements of **B** are represented by $B_{l,m}$ (B_1 to B_4) and $n_a = n_{i+1,k}$, $n_b = n_{i+1,k+1}$. B_0 *i.e.*, the elements corresponds to average coefficients can be given as $B_0 = \{1, 1, 1, 1, 1\}$.

Elements of **B** in the Equation 6 can be calculated using the following matrix:

$$\mathbf{B} = \begin{pmatrix} B_0 \\ B_1 \\ B_2 \\ B_3 \\ B_4 \end{pmatrix} = \begin{pmatrix} 1 & 0.82 & 0.91 & 1.58 & 0.00 \\ 1 & 0.82 & 0.91 & -1.58 & 0.00 \\ 1 & 0.82 & -1.82 & 0.00 & 0.00 \\ 1 & -1.22 & 0.00 & 0.00 & 1.58 \\ 1 & -1.22 & 0.00 & 0.00 & -1.58 \end{pmatrix} \tag{7}$$

This concludes our work on the wavelet transforms for the sound histogram. Subsequently, we will perform a dimension reduction technique for the wavelet transformed histograms sound samples using PCA.

3.2. Principle Component Analysis

Histogram sound samples provide compressed information about ambient sound intensity but also induces a high correlation between each bin. This causes instability in distance metric calculation when performing clustering. Moreover, we can simply prove that this high correlation is not affected by the WT by given the fact that it is a linear transformation (by using the fact $W[X] = aX + b$ then, $W[\bar{X}] = a\bar{X} + b$). The result of such is displayed in Fig. 2.

Fig. 2(a) illustrates the correlation between $d_{j,k}$, which is calculated for a given sensor node. It is observed that applying PCA, the high correlation between data has been
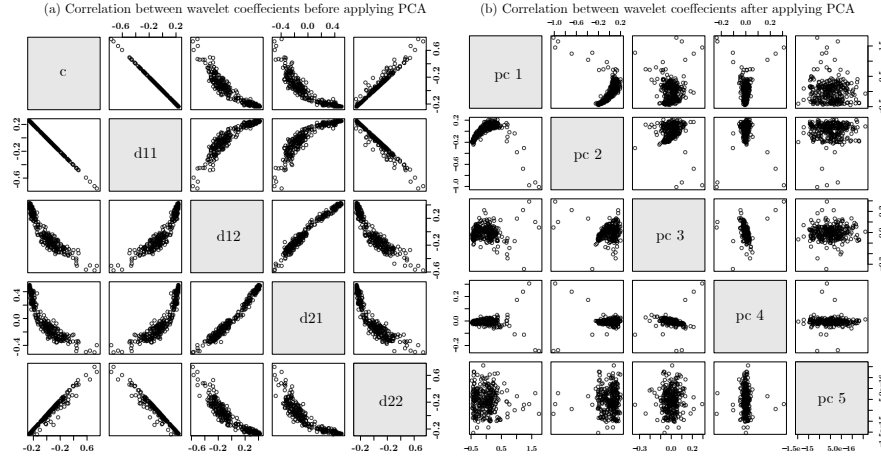
Figure 2: Comparing the correlation coefficient between wavelet coefficients (a) before PCA (b) after PCA.

removed. Note that, the column and row represent the coefficients and the scattered plot represents the correlation between two different coefficients. Thus, it is not hard to see that there is a high correlation between $d_{j,k}$. We transform the sound histogram into an uncorrelated space using PCA to remove the high correlation in between. Consequently, PCA provides an optimal solution for the following optimization problem:

$$
\begin{cases}
\max & e_i^T \Sigma e_i \\
\text{subject to} & e_i^T e_i = 1 \\
& e_j^T \Sigma e_i = 0, \text{ for } j < i,
\end{cases} \tag{8}
$$

where Σ is the covariance matrix of calculated $d_{j,k}$ for sound histogram (for one day of a given sensor node). e_i, e_j are the eigenvectors associated with Σ in the optimal solution, which only satisfies the Karush Kuhn Tucker (KKT) optimality conditions. We denote the sound histograms for a one day (288 represents data collected for 5-minute windows over 24 hours) single sensor node as $\mathbf{H} = \{\mathbf{w_1}, \mathbf{w_2}, \ldots, \mathbf{w_{288}}\}$ and $\mathbf{w}$ is the wavelet coefficient vector for each sound histogram. Afterward, the projection

10

of $\mathbf{H}$ onto *PC*s can be represented as follows:

$$\mathbf{Z} = \begin{pmatrix} e_1 \\ . \\ . \\ . \\ e_5 \end{pmatrix} \mathbf{H} \tag{9}$$

where $\mathbf{Z}$ is the projection of $\mathbf{H}$. Fig. 2 shows the transformation of sound histograms into uncorrelated space after applying PCA.

Let $\mathbf{U} = \{\mathbf{e_1}, ..., \mathbf{e_5}\}$ then the reconstructed sound histograms $\hat{\mathbf{H}}$ from *PC*s can be rewritten as shown below:

$$\hat{\mathbf{H}} = \mathbf{U}\mathbf{Z} \tag{10}$$

Mathematically $\hat{\mathbf{H}} = \mathbf{H}$ when we reconstruct the sound samples by including all the PCs. Using this method, a truncated reconstruction with only a selected subset of PCs would yield some components being discarded. This can be used to remove redundant information and hence a cleaner data. This method has been used in [27, 28], where redundant information is removed while preserving a relatively low variation of data.

As shown in [29, 30], there is a few formal ways of selecting the optimal subset of PCs $\{pc_1, ..., pc_k\}$. We select k by minimizing the projection error between $\hat{\mathbf{H}}$ and $\mathbf{H}$ given that $k < 5$. Fig. 3 (a) and Fig. 3 (b) illustrate the $\mathbf{H}$ and $\hat{\mathbf{H}}$ respectively. The reconstructed sound histograms $\hat{\mathbf{H}}$ are constructed using a subset of principal components from $\mathbf{Z}$. We select $k = 4$ (4 principal components) and the simple illustration of such can be found in Fig. 3. It can be observed that $\hat{\mathbf{H}}$ has successfully removed the low variational components across all five sound histogram bins when compared with the $\mathbf{H}$. By removing the low variational components, it also removes the redundant information inside the sound histograms. Therefore, $\hat{\mathbf{H}}$ can guarantee a more robust and accurate feature representation of sound histograms.

To validate the efficiency of the aforementioned data processing technique, we perform a simple hierarchical clustering method to show the difference between WT and PCA in terms of feature extraction based on sound histograms' similarity. Fig. 4 illustrates how the WT and the PCA enhance feature extraction based on the similarity of
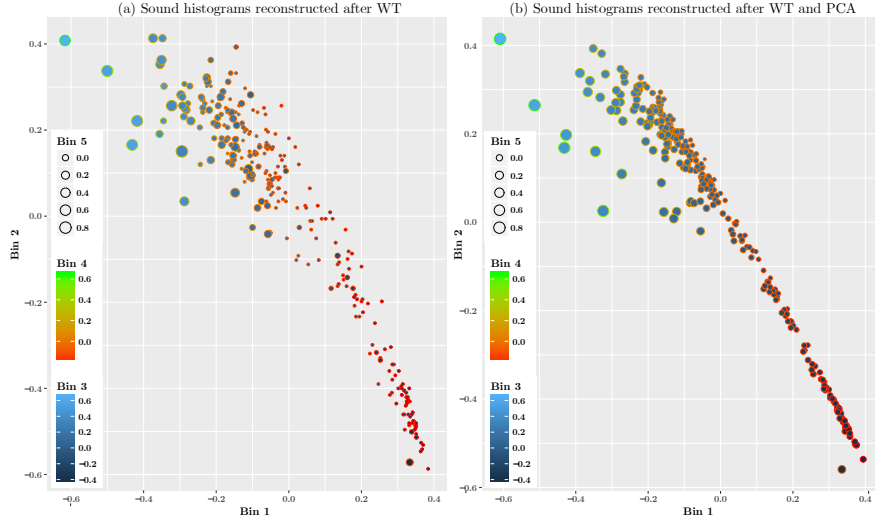
11

Figure 3: Comparison of sound data using different data processing techniques. The first two dimension of the histograms are shown in x and y axes, while magnitude of the 3rd dimension is given by the area of each circle. The 4th dimension is shown by the color intensity level of the perimeter of each circle and the 5th dimension is presented by the color intensity level of the area for each circle.

data.

Note that, the main focus here is to show the clustering result performed on different processing techniques rather than the technical detail of clustering (will be discussed in the next section). In addition, the markers presented in the figure do not yield any similarity between data processing despite the markers are using the same colors. It only represents the similarity corresponding to the exact data sample throughout the day, where it forms a different clustering group. Once again, this does not show the activities being detected but rather displays the group of similar processed sound samples. Fig. 4(a) shows that most data points are quite similar to each other for most of the time. Despite 4 clusters being generated, it does not yield any meaningful insight. In Fig. 4(b), we observe there is a different cluster that appears in the morning to the evening period. However, this still does not show the variation between morning and evening. Lastly, after performing PCA and WT as shown in Fig. 4(c), we notice there are variations in the morning, afternoon, and evening. But this requires further investigation, which we will discuss in the subsequent sub-section. Here, we show that
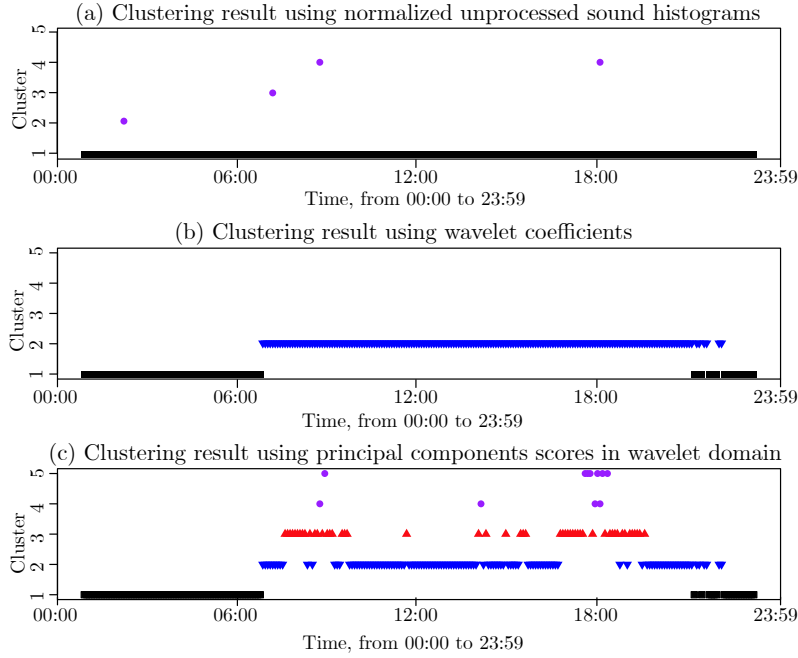
Figure 4: Clustering result using different preprocessing techniques: (a) using only normalized unprocessed sound histograms. (b) using wavelet coefficients. (c) using principal components scores in wavelet domain.

by applying WT and PCA techniques, it is possible to extract more variation in the histogram sound samples for further insights' extraction.

3.3. Sound Histograms Clustering

In this sub-section, we focus on the clustering technique for processed sound data using the aforementioned techniques. First, we calculate the average value for each 5-minute sound histogram, and we noticed there is a latent hierarchical structure in sound histograms despite variation of behavior for different days of the week (Weekday, Saturday, and Sunday). Therefore, it is crucial to trench this hidden structure in order to establish a model for daily usage using the agglomerative hierarchical clustering method [31].

Agglomerative hierarchical clustering method provides variation to the distance measure between two sound histograms samples, where the only measurement is needed to formulate the cluster. This simplifies the clustering process for the PCA and WT re-

(a) Original Metric Space   (b) Distortion at CCC=0.30   (c) Distortion at CCC=0.83
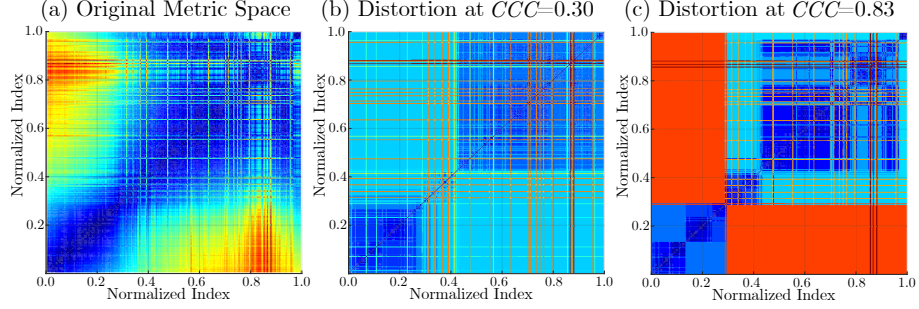
Figure 5: Comparison of distortion of metric space for sound histograms.

sults, while provide a visual clue to diagnose the clustering result. In our study, we generate distortion based on the distance between two sound histogram samples using Euclidean distance. We observe such behavior in sound histogram samples and it varies depending on different environmental conditions and time of the days. Therefore, it is crucial to find the correct similarity measurements for sound histograms on each day for each individual sensor nodes, so that we can reduce the distortion of the metric spaces. We calculate the Cophenetic Correlation Coefficient (CCC) as a goodness of fit static for the hierarchical dendrogram for each similarity measurement type. We study two different CCC values as illustrated in the Fig. 5.

The x and y axes denote the normalized index of the sound histogram, which can be calculated through min-max normalization of sound histogram w.r.t the number of sound histograms samples collected in a day. By visually observing the distortion of metric space, it is not hard to see that higher CCC value resembles the original metric space provides a more meaningful clustering outcome. Next, we choose the dendrogram that has the highest CCC value that shows the best clustering consistency as the foundation for clustering similar background noise.

4. Detection Phase

After defining the data preprocessing methods, we analyze the clustered data in order to detect events such as (1) raining period, (2) background noise, and (3) human outdoor activity.

14

4.1. Raining Period Estimation

Weather plays an important role in affecting human outdoor activity. Hence, it is critical to avoid including the raining period as part of human outdoor activity detection by estimating the raining period. To better estimate the raining period, multiple sensors (7 sensor nodes that are co-located in the same playground in our case) within the same PoI are used. The rationality of using multiple sensor nodes is that the rain is the common event that will be captured by all the sensors in the vicinity at the same time-line. To provide a comparison for raining period detection, resistive water droplet sensors are used for validation. The resistive water droplet sensors output a value ranging from 0 to 1023, where lower value denote presence of water droplet while higher value denote the opposite. Note that, the actual raining period detected by resistive water sensors can be shorter due to the remaining water droplet on the sensor surface.

We use the clustering method for 7 sensor nodes' sound histogram, and we observe that are cluster that resembles raining period compared to non-raining period. To further validate the accuracy of the raining period, we investigate the resistive raining sensor against the estimated raining period as shown in Fig. 6.

The solid line represents resistive rain value for each sensor and the diamond markers denote raining period detected using sound sensor. We notice there is a slight delay in identifying the raining period when compared it to the resistive rain sensor. However, in term of actual raining period, sound sensor can estimate the time of rain stop more accurate than the resistive round sensor. To better separate the raining period detection and human outdoor activity, prolong period of high ambient noise intensity is one of the significant characteristic. Any other short impulse of noise periods (e.g. heavy trucks, car, thunder, train, etc.) would be discarded when compared to the actual raining period. Thus, we show that the raining period estimation is possible using the clustering method given the interval of data samples is sufficient.

4.2. Background Noise Detection

Before identifying the human outdoor activity, we need to understand the characteristic of background noise period for each sensor node at the given PoI. First, we define
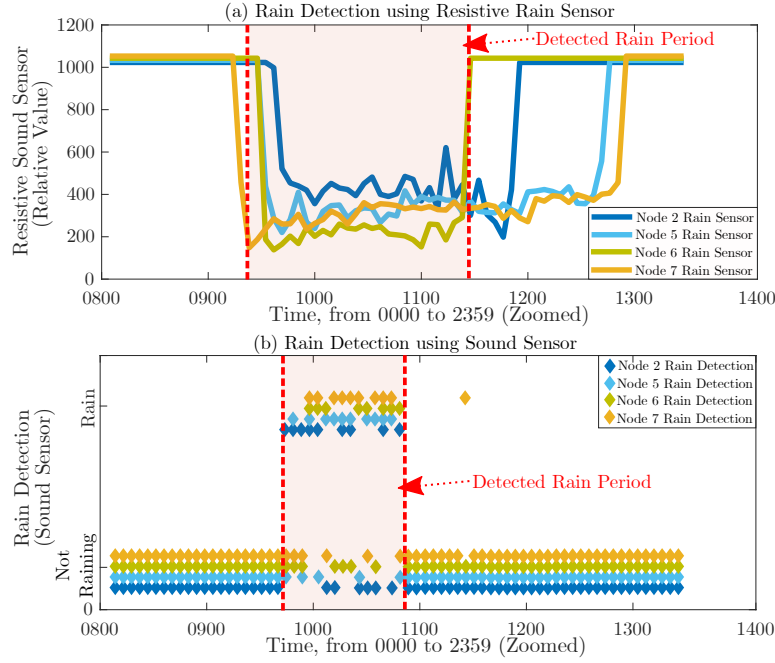
Figure 6: Comparing rain detection method using (a) rain sensor data (solid line) and (b) noise sensor data (diamond markers) based on the data collected on 28th July 2016.

the ambient background noise in our system model given the human activity is absent. The most straightforward approach to understand the background noise is to restrict any potential human activities for weeks to record a human activity free background noise, but this method would prevent the resident to utilize the PoI. Thus, we attempt to understand the background noise based on an assumption that the human activity does not occur at the same time throughout the week and months. Based on this assumption, we can detect the human activity using the outliers detection method (will be further elaborated in the next sub-section 4.3). In order to identify the background noise, we use the aforementioned clustering technique with the processed data (applying WT and PCA) for each day. An example of the background noise periods throughout different day of the week can be found in Fig. 7.

By observing the general patterns of the background period, we divide it into the following categories in Table 2 to match their characteristic:

(a) Background noise periods detection for Node 7

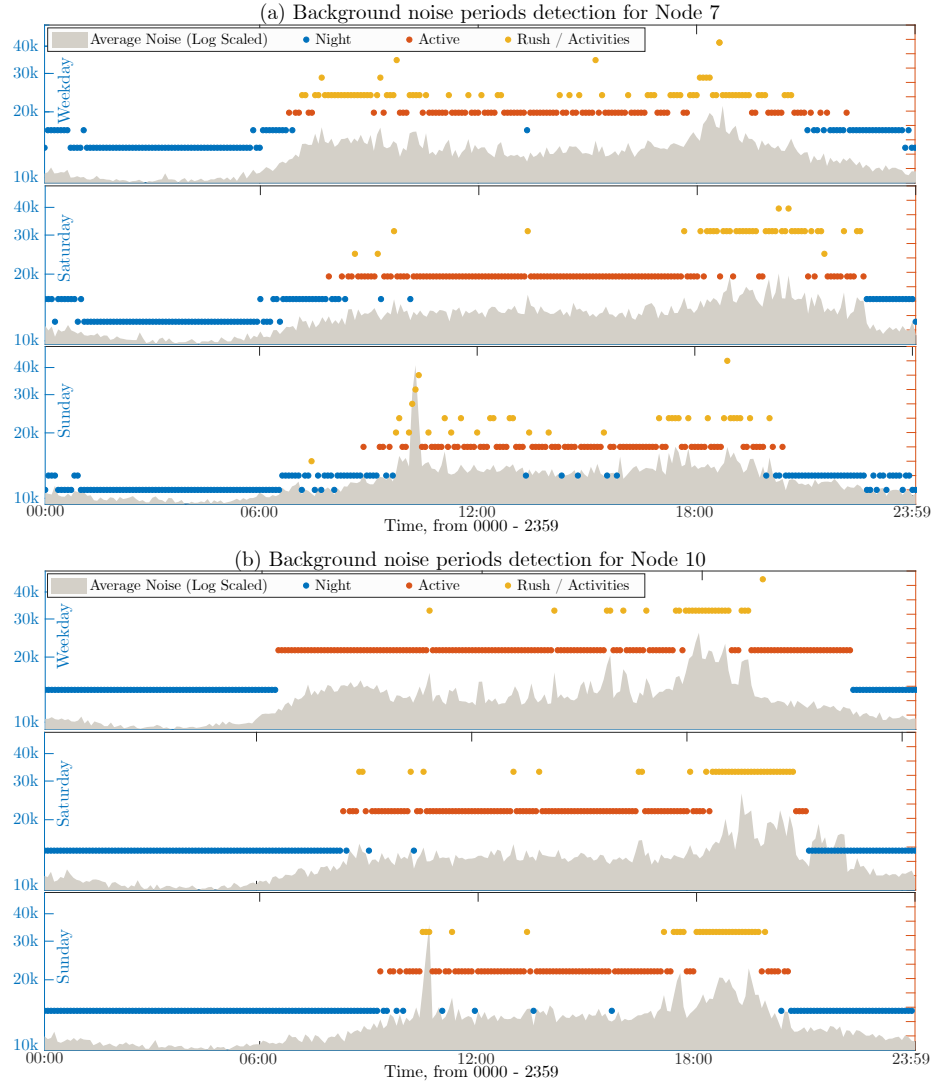(b) Background noise periods detection for Node 10

Figure 7: Background noise periods detection for sensor Node 7 and 10. (Top: Weekday 2nd August 2016. Middle: Saturday 16th July 2016. Bottom: Sunday 10th July 2016.)

Table 2: Background Period Definition and Description

Background Period	Description
Quiet/Night (NP)	There is minimal or no human outdoor activity going on.
Active (AP)	There is some outdoor activities going on.
Rush Hour period (RP)	There is large number of outdoor activities going on.

(a) Clusters Pattern of Node 7 over two months   (b) Clusters Pattern of Node 10 over two months
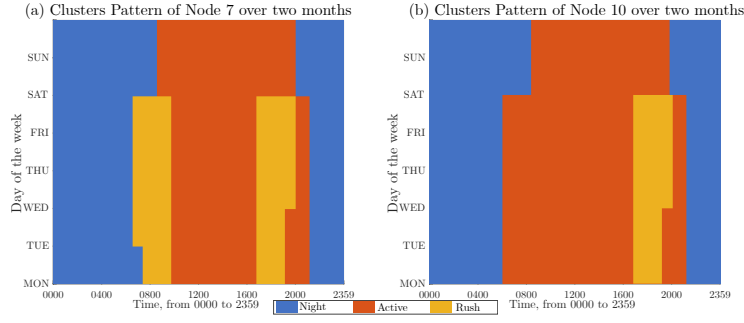
Figure 8: Average of periods pattern over a two months period (1st July 2016 - 31st August 2016) for Node 7 and 10.

After defining individual background noise periods, we observed some background noise behavior from the data shown in Fig. 8. For instance, there is long NP on weekend, which is possible due to the majority of the residents are not working at weekend. This also contributed to a shorter RP at weekend. In addition, there is an increment of the outdoor activities (scattered points) on Sunday compared to Saturday and weekday. In the Fig. 8(a), there is two obvious rush hour periods during morning and evening, where in Fig. 8(b), it only shows the evening rush hour period. The reason why sensor node 7 have two peaks detected is because it is located in the proximity of road, railway line, and pedestrian path. Meanwhile, sensor node 10 is located far away from these locales (roughly 400m apart) and capture less background noise compared to sensor node 7.

However, if we attempt to detect human outdoor activity only with the noise level, we are not able to differentiate the background noise and human outdoor activity. Instead, we try to understand the background noise across different time of the day and week to provide a noise pattern study before detecting any outdoor activity.

4.3. Outdoor Activity Detection

After identifying the background period for each sensor, we need to further investigate methods to distinguish an outdoor activity and background noise using outlier detection. Outdoor activity can be defined as the presence of human outdoor activity in a particular PoI. This includes kid's playing and large group chatting, which happens in the vicinity of the sensor node. As observed, the average value level of RP is much higher compared with AP and NP with different deviation values. In Fig. 7, we notice that the average noise value varies through different times of the day, and it is hard to draw a meaningful conclusion out of it. Therefore, a simple and robust method such as applying a threshold is not suitable under these constraints. In addition, the variation of background noise collected in each sound sensor is quite high, requiring different thresholding values and it is not efficient to manually devise a threshold for each sound sensor. The chances of This further increase of false-positive (falsely label background noise as outdoor activity) and false negative (falsely label outdoor activity as background noise) in the outdoor detection model. Therefore, data samples from different background noise periods need to be taken into consideration when identifying the outdoor activities separately.

In the outdoor activity detection module, we calculate a chi-square statistic for each sound histograms samples to distinguish background noise and human outdoor activity. Chi-square distribution with k degrees of freedom is described as a distribution of the squares for k independent standard normal random variables. The corresponding of the Chi-square distribution with k degrees of freedom can be calculated when optimal k PCs are achieved through independence between PCs. For example, we can show two PCs are independent by proving the product of two eigenvectors is zero by applying the symmertricity of two covariances as follows:

Suppose $\lambda_i \neq \lambda_j$ and $\Sigma e_i = \lambda_j e_i$ and $\Sigma e_j = \lambda_j e_j$ where λ_i, λ_j are eigenvectors

19

of the symmetric matrix Σ, with corresponding eigenvectors, e_i and e_j respectively. Then We have the following:

$$(\lambda_i - \lambda_j)(e_i - e_j) = (\lambda_i e_i, e_j) - (e_i, \lambda_j e_j)$$
$$= (\Sigma e_i, e_j) - (e_i, \Sigma e_j)$$

By using the symmetricity of Σ we can rewrite the equation as below:

$$(\Sigma e_i, e_j) - (e_i, \Sigma e_j) = \Sigma e_i . e_j - e_i . \Sigma e_j$$
$$= e_i \Sigma^T e_j - e_i . \Sigma e_j$$
$$= e_i \Sigma e_j - e_i . \Sigma e_j = 0$$

Since $\lambda_i \neq \lambda_j$, it shows that product of any two eigenvectors of a symmetric matrix is zero.

After denoting zero symmetric matrix and optimality of k PCs, we compute the corresponding chi-square statistic $\chi^2_{(i)}$ of k degrees of freedom for sound histogram samples as follows:

$$\chi^2_{(i)} = pc^2_{(1,i)} + pc^2{}_{(2,i)} + ... + pc^2_{(k,i)}, \tag{11}$$

where $pc^2_{(n,i)}$ is the standardized PC score of the ith sound histogram (all sound histograms correspond to the same background noise period) and $n \in \{1, ..., k\}$.

If the background noise remains stationary (sound histograms within the same cluster yields similar property) between sound histograms corresponding to the same time of the day, we can denote that particular data sample has a lower probability of being classified as outdoor activities. The intuition behind is that a sound histogram, which possesses lower probability values represents the outdoor activities that deviate significantly from a background noise period. Using the aforementioned concept, we can further formulate a binary output for activity detection using the following equation:

$$h_{activity} = \begin{cases} 1 & \text{if} \quad \chi^2_{(i)} \geq \beta \\ 0 & \text{if} \quad \chi^2_{(i)} < \beta \end{cases} \tag{12}$$

where β is a critical value in k degrees of freedom chi-square distribution and $h_{activity}$ is a binary value that represents the outdoor activity for a given sound histogram samples. In addition, $\chi^2_{(i)}$ denotes the activity level for each sound histogram sample. As
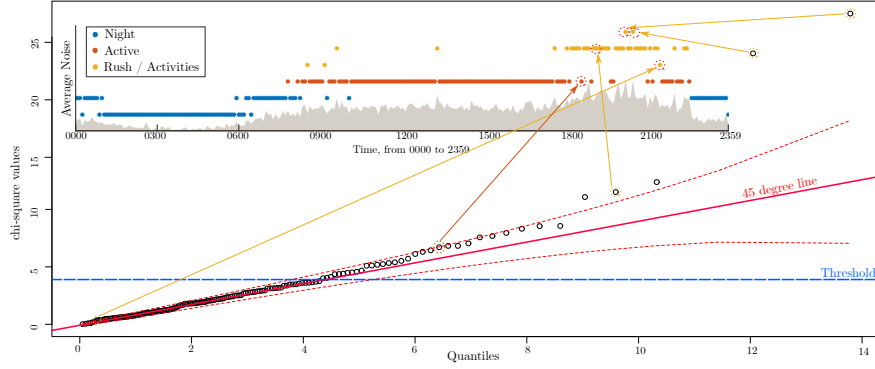
Figure 9: Quantile-Quantile plot of the chi-square distribution for the active period on Saturday 16th July 2016 (Node 7)

a remark of the outdoor activity detection, the period detected as raining would be labeled as zero outdoor activities, so that both events do not overlap with each other.

Next, we need to determine the value of β, which it is required to define the lower probability separation of the sound histograms sample into outdoor activity and background noise. We devise the approach of using quantile-quantile (q-q) plot [32, 33, 34] of a background noise period to visually identify the deviation sound sample at a given day of study. In the q-q plot, computed chi-square quantiles are plotted against the theoretical chi-square quantiles, and we can assess whether our computed values plausibly come from the same theoretical distribution. If the value does not fall in the same theoretical distribution, it implies that particular interval has a higher probability of outdoor activity.

We chose a Saturday (16th July 2016) as one of the days to study outdoor activity using sensor node 7. The q-q plot of the noise samples is illustrated in Fig. 9. From the q-q plot, we observe chi-square values begin to deviate significantly from the $\angle 45$ line beyond a certain threshold T. Based on the patterns, we observe some sound samples are related to a different distribution compared to a chi-square distribution. For instance, the period of 18:00 to 21:00 is supposed to be a rush period, but one point that is highlighted by the orange arrow shows an active period instead. These data point that falls outside the red dotted line, it is considered as outliers and it has a higher probability of an outdoor activity being detected. Meanwhile, those located within the

21

red dotted line or close to $45 \deg$ line is either background noise or lower probabilities labeled as outdoor activities. By tracing back the sound histograms temporal element, we can obtain the occurrence time for data samples and reconstruct the time series data with a label for each sound histogram samples. Note that, the same procedure can be repeated for other sensor nodes and obtain a similar result.

Apart from the q-q plot approach, T also can be estimated from other change point estimation methods such as the hybrid approach [35], Bayesian change point estimation [36], and wavelet methods [37], etc. Nonetheless, we use the $T = 0.43$ as β for the outdoor detection method for our system model. Thus, this concludes the outdoor activity detection module, and we will perform validation in the next section.

5. Evaluation

In this section, we validate the detected events by the proposed method against the ground truth to verify the accuracy of the proposed outdoor activity detection model. The same open space playground located near the residential area studied in the previous section is also used to collect ground-truth data. There are a total of 7 sensor nodes installed throughout the playground area (1650m^2), which is also next to the metropolitan railway line, bus-stop, and main road. These combinations of nearby amenities produce a PoI with noisy ambient noise varying from time to time, which is an ideal test case to prove the effectiveness of our model. We collect the ground-truth data on a Saturday evening (20th July 2017) for 5 hours starting from 14:00 to 19:00, where there is a high probability of human outdoor activities. Although there are 7 sensor nodes located at the PoI, we only focus on 4 sensor nodes around the premises that are within our range of sight as shown in Fig. 10(2) with the magenta arrow. We captured the video for 5 hours and invited 3 volunteers to label the outdoor activity time stamp in the video as ground-truth data. Throughout the data collection, there are roughly two hours (16:00 - 18:20) of activities being recorded as shown in Fig. 10.

Using a similar threshold from the previous section, we identified the outdoor activity and compared them to the ground-truth. The aggregation of 4 sensors is achieved through a "or" function, where one sensor detected outdoor activity represents the pres-
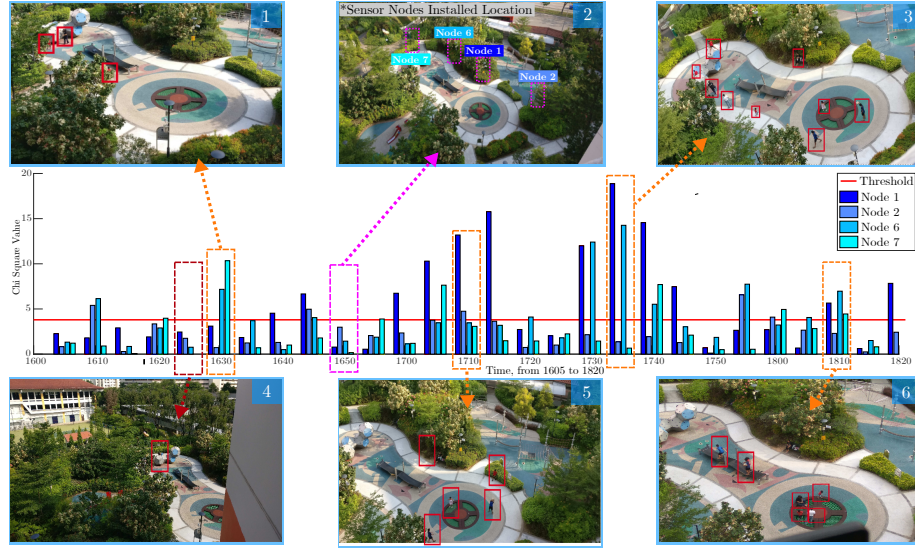
Figure 10: Obtained chi-square values with the ground truth images (snipped from the recorded video) for each sensor nodes on 20th July 2017 evening (only four out of seven sensor nodes are presented here).

ence of human outdoor activity in that PoI at that particular timestamp. After comparing the computed result and ground-truth data, it is not hard to see that the number of people in each image closely resembles the aggregated output of the sensor nodes.

Table 3: Experimental Results

Methods	True Detected	False Positive	False Negative
Raw data	65.0%	35.0%	0.0%
Raw data + WT	60.2%	34.3%	5.5%
Raw data + PCA	50.0%	33.5%	16.5%
Raw data + PCA + WT	85.8%	3.5%	10.7%

Next, we generate the confusion matrix based on the comparison of different processing methods and ground-truth. The result is shown in Table 3. It can be observed that approaches without applying WT and PCA (raw data only, raw data and PCA as

23

well as raw data and WT) yields higher false positive, which contributes to lower true detected values (65.0%, 60.2%, and 50% respectively). To understand the reason behind false-negative rates of 10.7% for proposed approach using PCA and WT, we study the ground-truth at the corresponding timestamps. It reveals that these false-negative events corresponded when there are less than one or two people within the PoI that is quiet. An example of the false-negative event can be found in Fig. 10(4). It shows there is only one person in the PoI at that time-stamp and remains quiet and hence not adequate sound information to capture the event accurately. As for other occupancy detection, it remains quite accurate most of the time. Thus, we can draw a conclusion that the system model we proposed work effectively and requires less human effort to provide training data.

6. Conclusion

In this paper, we have shown how the sound histogram can be used as a data reduction method to collect insights about the residential area rather than sending data samples frequently. Using methods such as WT and PCA, we demonstrate it is possible to generate more robust and meaningful feature sets before performing a clustering algorithm to group similar data patterns. In order to further extract insights from the sound histogram data, we devise a system model to detect background noise period, human outdoor activity, and raining period. Lastly, we validate our model against ground-truth data and managed to identify human outdoor activity with an accuracy of 85.5% using only sound sensor.

In future works, we expect the system model to adapt different sources of data to enrich the outdoor activity detection patterns. Multiple sensors will be used as a foundation of the sensor fusion to generate more accurate data is also part of the study. Moreover, the trade-off between normal sampling rate versus sound histograms could be further investigated from a hardware or information theory perspective is quite interesting. Also, the optimal histogram bin size and range choices also can be throughout investigated in order to increase detection accuracy. This will provide us a more optimal system to design data collection platforms from different aspects such as energy

consumption of sensor nodes, communication overhead, and data quality.

Acknowledgment

References

[1] I. Katakis, Mining urban data (part a), Information Systems 54 (2015) 113 – 114. doi:https://doi.org/10.1016/j.is.2015.08.002.
URL http://www.sciencedirect.com/science/article/pii/S0306437915001374

[2] B. P. L. Lau, N. Wijerathne, B. K. K. Ng, C. Yuen, Sensor fusion for public space utilization monitoring in a smart city, IEEE Internet of Things Journal 5 (2) (2018) 473–481.

[3] B. P. L. Lau, S. H. Marakkalage, Y. Zhou, N. U. Hassan, C. Yuen, M. Zhang, U.-X. Tan, A survey of data fusion in smart city applications, Information Fusion 52 (2019) 357 – 374. doi:https://doi.org/10.1016/j.inffus.2019.05.004.
URL http://www.sciencedirect.com/science/article/pii/S1566253519300326

[4] L. Jayasinghe, N. Wijerathne, C. Yuen, M. Zhang, Feature learning and analysis for cleanliness classification in restrooms, IEEE Access 7 (2019) 14871–14882.

[5] N. Wijerathne, S. K. Viswanath, M. S. Hasala, V. Beltran, C. Yuen, H. B. Lim, Towards comfortable cycling: A practical approach to monitor the conditions in cycling paths, in: 2018 IEEE 4th World Forum on Internet of Things (WF-IoT), 2018, pp. 778–783. doi:10.1109/WF-IoT.2018.8355173.

[6] B. P. L. Lau, T. Chaturvedi, B. K. K. Ng, K. Li, M. S. Hasala, C. Yuen, Spatial and temporal analysis of urban space utilization with renewable wireless sensor network, in: 2016 IEEE/ACM 3rd International Conference on Big Data Computing, Applications and Technologies, ACM, 2016, pp. 133–142.

[7] A. Tyndall, R. Cardell-Oliver, A. Keating, Occupancy estimation using a low-pixel count thermal imager, IEEE Sensors Journal 16 (10) (2016) 3784–3791.

[8] Z. Liu, J. Huang, J. Han, S. Bu, J. Lv, Human motion tracking by multiple rgbd cameras, IEEE Transactions on Circuits and Systems for Video Technology 27 (9) (2017) 2014–2027. `doi:10.1109/TCSVT.2016.2564878`.

[9] T. Yu, J. Zhao, Y. Huang, Y. Li, Y. Liu, Towards robust and accurate single-view fast human motion capture, IEEE Access 7 (2019) 85548–85559. `doi:10.1109/ACCESS.2019.2920633`.

[10] Q. Hao, F. Hu, Y. Xiao, Multiple human tracking and identification with wireless distributed pyroelectric sensor systems, IEEE Systems Journal 3 (4) (2009) 428–439. `doi:10.1109/JSYST.2009.2035734`.

[11] A. Booranawong, N. Jindapetch, H. Saito, Adaptive filtering methods for rssi signals in a device-free human detection and tracking system, IEEE Systems Journal 13 (3) (2019) 2998–3009. `doi:10.1109/JSYST.2019.2919642`.

[12] Y. Zhou, B. P. L. Lau, Z. Koh, C. Yuen, B. K. K. Ng, Understanding crowd behaviors in a social event by passive wifi sensing and data mining, IEEE Internet of Things Journal (2020) 1–1`doi:10.1109/JIOT.2020.2972062`.

[13] K. Li, C. Yuen, S. S. Kanhere, K. Hu, W. Zhang, F. Jiang, X. Liu, An experimental study for tracking crowd in smart cities, IEEE Systems Journal 13 (3) (2019) 2966–2977. `doi:10.1109/JSYST.2018.2880028`.

[14] G. Bieber, A. Luthardt, C. Peter, B. Urban, The hearing trousers pocket: Activity recognition by alternative sensors, in: Proceedings of the 4th International Conference on PErvasive Technologies Related to Assistive Environments, PETRA '11, Association for Computing Machinery, New York, NY, USA, 2011.

doi:10.1145/2141622.2141674.

URL https://doi.org/10.1145/2141622.2141674

[15] A. Longo, M. Zappatore, M. Bochicchio, S. B. Navathe, Crowd-sourced data collection for urban monitoring via mobile sensors, ACM Trans. Internet Technol. 18 (1) (Oct. 2017). doi:10.1145/3093895.

URL https://doi.org/10.1145/3093895

[16] A. J. Fairbrass, M. Firman, C. Williams, G. J. Brostow, H. Titheridge, K. E. Jones, Citynet—deep learning tools for urban ecoacoustic assessment, Methods in Ecology and Evolution 10 (2) (2019) 186–197. arXiv:https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.13114, doi:10.1111/2041-210X.13114.

URL https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13114

[17] R. Kitchin, Getting smarter about smart cities: Improving data privacy and data security, Data Protection Unit, Department of the Taoiseach, Dublin, Ireland. (2016).

[18] A. Bartoli, J. Hernández-Serrano, M. Soriano, M. Dohler, A. Kountouris, D. Barthel, Security and privacy in your smart city, in: Proceedings of the Barcelona smart cities congress, Vol. 292, 2011, pp. 1–3.

[19] S. Rekik, D. Guerchi, S.-A. Selouani, H. Hamam, Speech steganography using wavelet and fourier transforms, EURASIP Journal on Audio, Speech, and Music Processing 2012 (1) (2012) 20.

[20] C. Bilen, S. Huzurbazar, Wavelet-based detection of outliers in time series, Journal of Computational and Graphical Statistics 11 (2) (2002) 311–327. arXiv:https://doi.org/10.1198/106186002760180536, doi:10.1198/106186002760180536.

URL https://doi.org/10.1198/106186002760180536

[21] M. K. H. Kolekar, G. L. Raja, S. Sengupta, An introduction to wavelet-based image processing and its applications, in: Computer Vision: Concepts, Methodologies, Tools, and Applications, IGI Global, 2018, pp. 110–128.

[22] J. Lever, M. Krzywinski, N. Altman, Points of Significance - Principal Component Analysis, Nature Methods 14 (7) (2017) 641–642. `doi:10.1038/nmeth.4346`.
URL `https://doi.org/10.1038/nmeth.4346`

[23] J. Han, J. Pei, M. Kamber, Data mining: concepts and techniques, Elsevier, 2011.

[24] F. Abegaz, K. Chaichoompu, E. Génin, D. W. Fardo, I. R. König, J. M. Mahachie John, K. Van Steen, Principals about principal components in statistical genetics, Briefings in Bioinformatics 20 (6) (2018) 2200–2216. `arXiv:https://academic.oup.com/bib/article-pdf/20/6/2200/31789463/bby081.pdf,doi:10.1093/bib/bby081`.
URL `https://doi.org/10.1093/bib/bby081`

[25] C. M. Leavey, M. N. James, J. Summerscales, R. Sutton, An introduction to wavelet transforms: a tutorial approach, Insight-Non-Destructive Testing and Condition Monitoring 45 (5) (2003) 344–353.

[26] T. K. Sarkar, C. Su, R. Adve, M. Salazar-Palma, L. Garcia-Castillo, R. R. Boix, A tutorial on wavelets from an electrical engineering perspective. i. discrete wavelet techniques, IEEE Antennas and Propagation Magazine 40 (5) (1998) 49–68. `doi:10.1109/74.735965`.

[27] L. Zhang, R. Lukac, X. Wu, D. Zhang, Pca-based spatially adaptive denoising of cfa images for single-sensor digital cameras, IEEE Transactions on Image Processing 18 (4) (2009) 797–812. `doi:10.1109/TIP.2008.2011384`.

[28] T. Jagadesh, R. J. Rani, A novel speckle noise reduction in biomedical images using pca and wavelet transform, in: 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 2016, pp. 1335–1340. `doi:10.1109/WiSPNET.2016.7566354`.

[29] R. B. Cattell, The scree test for the number of factors, Multivariate Behavioral Research 1 (2) (1966) 245–276, pMID: 26828106. `arXiv:https://doi.org/10.1207/s15327906mbr0102_10`, `doi:10.1207/s15327906mbr0102\_10`.
URL `https://doi.org/10.1207/s15327906mbr0102_10`

[30] M. A. Carreira-Perpinán, A review of dimension reduction techniques, Department of Computer Science. University of Sheffield. Tech. Rep. CS-96-09 9 (1997) 1–69.

[31] F. Murtagh, P. Contreras, Algorithms for hierarchical clustering: an overview, WIREs Data Mining and Knowledge Discovery 2 (1) (2012) 86–97. `arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/widm.53`, `doi:10.1002/widm.53`.
URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/widm.53`

[32] J. D. Gibbons, S. Chakraborti, Nonparametric Statistical Inference: Revised and Expanded, CRC press, 2014.

[33] W. Widhiarso, B. Sumintono, Examining response aberrance as a cause of outliers in statistical analysis, Personality and Individual Differences 98 (2016) 11 – 15. `doi:https://doi.org/10.1016/j.paid.2016.03.099`.
URL `http://www.sciencedirect.com/science/article/pii/S0191886916302525`

[34] C. Leys, O. Klein, Y. Dominicy, C. Ley, Detecting multivariate outliers: Use a robust variant of the mahalanobis distance, Journal of Experimental Social Psychology 74 (2018) 150 – 156. `doi:https://doi.org/10.1016/j.jesp.2017.09.011`.
URL `http://www.sciencedirect.com/science/article/pii/S0022103117302123`

[35] S. Boutoille, S. Reboul, M. Benjelloun, A hybrid fusion system applied to off-line detection and change-points estimation, Information Fusion 11 (4) (2010) 325 –

337. doi:https://doi.org/10.1016/j.inffus.2009.10.004.
URL http://www.sciencedirect.com/science/article/pii/
S1566253509000918

[36] S. Aminikhanghahi, D. J. Cook, A survey of methods for time series change point detection, Knowledge and information systems 51 (2) (2017) 339–367.

[37] T. Ogden, E. Parzen, Change-point approach to data analytic wavelet thresholding, Statistics and Computing 6 (2) (1996) 93–99. doi:https://doi.org/10.1007/BF00162519.