Detecting Smoothness of Pedestrian Flows by Participatory Sensing with Mobile Phones

Tomohiro Nishimura Takamasa Higuchi Hirozumi Yamaguchi Teruo Higashino
Graduate School of Information Science and Technology, Osaka University
1-5 Yamadaoka, Suita, Osaka 565-0871, Japan
{t-nisimr, t-higuti, h-yamagu, higashino}@ist.osaka-u.ac.jp

ABSTRACT
Crowd density and behavior of pedestrian flows at each area in public space (*e.g.*, in a complex commercial building and in a large event site) would be essential information for advanced pedestrian navigation, early detection of crowd accidents and planning of evacuation guidance strategy for disaster control. In this paper, we propose a novel system for estimating crowd density and smoothness of pedestrian flows in public space based on participatory sensing by a small proportion of mobile phone users in the crowd. By analyzing walking motion of the pedestrians and ambient sound in the environment that can be monitored by accelerometers and microphones in off-the-shelf mobile phones, our system classifies the levels of congestion and behavior of pedestrian flows at each area into four categories that well represent the crowd behavior. Through field experiments using Android smartphones, we show that our system can recognize the current situation in the environment with accuracy of 60%–78%.

Author Keywords
Participatory sensing; crowd density estimation; pedestrian flows; ease of walking

ACM Classification Keywords
I.5.4. Pattern Recognition: Applications

INTRODUCTION
Pedestrian navigation is one of the most primary location-based services that billions of people use on a daily basis. While their basic function is to guide users to a specified destination along the shortest-distance route, some recent commercial mobile applications provide more advanced options [8]: In a rainy day, the system can recommend a route that goes through buildings, covered-in pathways and underground shopping areas, so that the user can get to the destination without using an unbrella. For users traveling with a large amount of baggage, it can also suggest a route without stairs. Thus, by fully utilizing map information, they effectively optimize their services according to the current context of the user, and provide the best route that the user can travel easily and comfortably.

In crowded public space like a complex commercial buiding and a large event place, travel time and ease of walking on the way to the destination would significantly depend on congestion along the path, as well as the total travel distance. Especially for the people with a disability and those who are accompanied by babies and infants, it would be preferable to avoid extremely crowded areas, where they are likely at risk of falls or may suffer from stress. Such congestion information has been successfully utilized for vehicle navigation systems, where measurement data collected from road-side units and location information of floating cars are analyzed online to minimize the total travel time [7, 12]. In contrast, to the best of our knowledge, none of the existing pedestrian navigation systems are aware of congestion along the paths in the route decision process. In addition to the navigation purposes, the information on crowd density and smoothness of pedestrian flows in public space would be also helpful for early detection and prevention of crowd accidents and planning of evacuation guidance strategy for disaster control.

Towards accurate crowd tracking in public space (*e.g.*, in a shopping mall), vision-based tracking techniques have been hevily investigated in the computer vision community [3]. Although such vision-based systems can reliably detect crowd density using existing infrastructure (*e.g.*, CCTV cameras), they usually have severe limitation on spatial coverage due to the limited viewing angle of ordinary camera devices.

To achieve sufficient coverage with less dependence on infrastructure, the idea of participatory sensing would be a strong alternative solution. One of the most simple approach would be to collect location information of mobile phones (that can be obtained by Wi-Fi fingerprinting [2, 15] or pedestrian dead reckoning (PDR) [5, 13]) to a centralized server. Analyzing these data at server-side, the system can roughly estimate spatial distribution of the pedestrians over the whole target area. However, such a relative density distribution would not offer sufficient information for estimating absolute crowd density in each region. Some recent work copes with the problem by employing short-range wireless communication via Bluetooth [14]. Mobile phones periodically probe neighboring devices, and then the number of detected neigh-

bors and temporal variance in received signal strength (RSS) are collected to a server to classify the crowd density into 7 categories. Although it can recognize the category of crowd density (*i.e.*, the approximate number of pedestrians per unit area) with accuracy of 41–88%, it is not aware of smoothness of the pedestrian flows which dominates ease of walking.

In this paper, we propose a novel system for estimating levels of congestion and smoothness of pedestrian flows at each region in public space. We assume that a small proportion of mobile phone users in a crowd contribute to the sensing service, and locally recognize the behavior of the surrounding pedestrians based on measurements from built-in sensors in their phones. In traffic engineering, it is known that walking motion of pedestrians exhibits different characteristics according to levels of congestion [9]: In crowded pathways, walking speed of each person is affected by that of the surrounding pedestrians, making intervals between their walking steps longer than his/her usual walking motion. In addition, at intersections of multiple pedestrian flows, people often dodge the surrounding pedestrians who walk toward a different direction. We capture such characteristic walking behavior in crowded situations by using accelerometers in mobile phones. In addition, due to crowd noise and convesation voice by the surrounding people, ambient audio would also have characteristic features in highly crowded areas. Such features can be captured by recording the surrounding ambient sound with microphones in the mobile phones. By extracting the motion-based features and the audio-based features from the accelerometer readings and audio recordings, respectively, each phone classifies the behavior of surrounding crowd into four categories: *low density*, *medium density*, *high density with smooth flows* and *high density with intersections*. The estimation results by each phone are associated with the phone's location (based on Wi-Fi fingerprinting or PDR) and are collected to a centralized server. By integrating these data from multiple phones, the system can estimate the behavior of pedestrian flows at each region.

We have implemented the sensing service on the Android platform, and have conducted a field experiment in an underground shopping area near a primary subway station in Osaka which accomodates more than 400,000 passernges per day. The experimental results show that our system could recognize the current situation at each area with accuracy of 60%–78%.

RELATED WORK AND CONTRIBUTION

A more cost-efficient approach to crowd density estimation is to utilize sensors in commercial mobile devices (*e.g.*, smartphones). Kannan et al. [4] count the number of mobile phone users in a crowd by exchanging audio beacons between the neighboring phones. The audio beacons contain frequency components that are associated with the phone's own ID (*e.g.*, MAC address) and the IDs of detected neighboring phones. By repeating such beacon exchange until the set of detected neighbors converges, the phones can recognize the number of phones in the crowd. Although it can count up to hundreds of phones by carefully designing the coding algorithm for the audio beacons, it can recognize only the mobile phone users who participate in the crowd counting service, keeping their speakers and microphones on. Weppner et al. [14] employ short-range wireless ad-hoc communication via Bluetooth for participatory crowd density sensing with mobile phones. To mitigate dependence on the proportion of mobile phone users who enable Bluetooth of their phones, they use variance in RSS values as well as the number of detected neighboring phones to classify the current crowd density in the target area into 7 categories. However, accuracy of crowd density estimation still significantly depends on the ratio of Bluetooth-enabled phones. Both of these systems focus on estimating the number or density of the crowd and do not care about the smoothness of the crowd flows, which would also be an important factor for pedestrian navigation and other crowd sensing applications.

Since ambient audio faithfully reflects the current situation in the environment, a microphone has been considered a powerful tool for recognizing the current context and surrouding situation of mobile phone users. Ear-phone [10] collects sound-levels at each location in metropolitan areas via participatory sensing with mobile phones, and constructs a noise map which can be utilized for city planning and pricing of real estate. It is relevant to our work in the sense that it also recognizes situation in the environment by analyzing magnitude of ambient audio. However, our goal is to recognize the behavior of the crowd based on accelerometer readings and audio recordings. For that purpose, we investigate the frequency components that well reflect the congestion level, and design an algorithm to recognize the current situation. Thus our work is substantially different from Ear-phone both in terms of the goal and the approach. SoundSense [6] utilizes the ambient audio for sound-based context recognition. Analyzing the audio that are recorded with mobile phones, it classifies types of the sound (*e.g.*, music, speech, etc.) and provides audio-based event detection (*e.g.*, walking, driving, etc.). Considering the limitation on computation power of commercial mobile devices, our system employs a simple machine learning algorithm for the sound-based congestion estimation. However, it would be worth mentioning that we may achieve finer-grained congestion estimation by effectively combining such event detection and audio fingerprinting techniques.

Our system monitors walking motion and ambient sound noise with mutiple sensors in off-the-shelf mobile devices, and collects these information to a server to estimate the current behavior of pedestrian flows. By analyzing walking motion of the pedestrians, it enables to estimate ease of walking at each locations in the target environment, which has not been considered in the existing crowd density sensing systems as in [14]. In addition, our system does not require wireless communication with the neighboring mobile devices, and reliably works even if the ratio of mobile phone users who participate in the sensing service is extremely low. To the best of our knowledge, this is the first system that simultaneously recognizes crowd density and smoothness of pedestrian flows using only sensor data from off-the-shelf mobile devices, without depending on manual reports by the users (*i.e.*, crowdsourcing). Furthermore, our system collects the

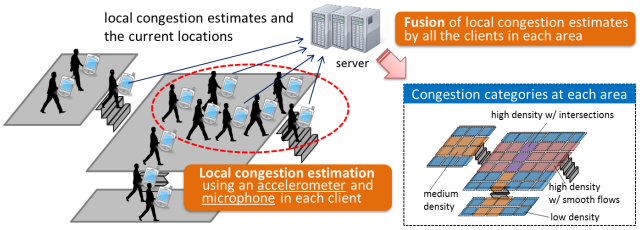


Figure 1. Architecture of the proposed system

sensing results from multiple mobile devices, and integrates these data based on a majority-vote algorithm. It reduces the impact of difference in walking motion among individuals, and effectively improves the estimation accuracy. This is also an important difference from the existing work in [14].

OVERVIEW

System Architecture

Figure 1 shows architecture of the proposed system. Our system is composed of mobile phones that are running a sensing application (*i.e.*, clients) and a server on a mobile cloud. Each client obtains accelerometer readings every 20 milliseconds, and analyzes the recent history of acceleration data to classify the current behavior of the pedestrian flow around the client into three categories: *low/middle density*, *high density with smooth flows* and *high density with intersections*. In addition, the clients also record the surrounding ambient sound with microphones in mobile devices, and apply spectrum analysis every 20 milliseconds. By analyzing the recent history of the audio spectrum, it classifies the crowd density of the surrounding enviroment into *low density*, *medium density* and *high density*.

Using the accelerometer readings and measurements from compass and gyro sensors, the clients also estimate its walking trajectories with pedestrian dead reckoning (PDR). By combining the estimated trajectories and the location information obtained by Wi-Fi fingerprinting as in [2, 15], the clients can estimate their current position with accuracy of a few meters. Then at each time slot t, a client i reports the current location $\boldsymbol{p}_{i,t}$, the acceleration-based congestion category $a_{i,t}$ and the sound-based congestion category $s_{i,t}$ to a server via cellular or Wi-Fi networks.

The server associates the estimation results $a_{i,t}$ and $s_{i,t}$ with pre-defined *areas* based on the location information $\boldsymbol{p}_{i,t}$, and integrates all the estimated congestion categories collected from the clients in each area. The areas are manually defined by the service provider, so that their size is sufficiently larger than the granularity of the location information provided by the clients. Behavior of pedestrian flows usually change at the points where width of the pathway becomes narrower, or at the locations with barriers or stairs. To achieve higher reliability, the areas should be determined by carefully considering structure of the building, so that the density and behavior of the crowd in the area is expected to have small variance.

The estimation results by each client may contain errors due to sensor noise and difference in walking motion among individuals. The server effectively mitigates the impact of such noise and provides reliable estimation of the bahavior of pedestrian flows by integrating the estimated congestion categories collected from each client.

Finally, the acceleration-based categories and the sound-based categories are integrated to classify the current situation at each region into four categories: *low density*, *medium density*, *high density with smooth flows* and *high density with intersections*.

The results can be utilized for estimating travel time to a destination, pedestrian navigation and visualization of congestion over the whole target area (*e.g.*, a shopping mall).

Definition of Congestion Categories

Older et al. [9] analyze the relationship between crowd density and walking speed of pedestrians in the crowd (see Figure 3). When the crowd density is less than 1.0 persons/m^2, average speed of the pedestrians keeps 75-80% of their original speed with 0.0 persons/m^2 (in case that each pedestrian walk alone). In contrast, the average speed steeply declines when the crowd density exceeds 1.0 persons/m^2 and becomes 20% of the original speed when the density is 2.5 persons/m^2. They also mention that if the crowd density exceeds 2.5 persons/m^2, there arises the risk of a serious accident where a number of pedestrians fall down one after another.

Based on the observations above, we classify the crowd density into the following three categories that are separated by the number of pedestrians per unit area:

- *Low density*: The situations where the crowd density is less than 1.0 persons/m^2 (Figure 2 (a)).
- *Medium density*: The situations where the crowd density is between 1.0 persons/m^2 and 2.5 persons/m^2 (Figure 2 (b))
- *High density*: The situations where the crowd density is more than 2.5 persons/m^2 (Figure 2 (c))

Another literature from architectural engineering [11] investigates physical and mental stress that arises when walking in a pedestrian flow. According to their report, the stress significantly increases when the people walk at intersections of multiple pedestrian flows. In order to suggest the safe and comfortable route, the number of locations where pedestrian flows intersect each other should be minimized.

From this point of view, we decided to divide the *high density* into the following two subcategories based on difference in walking directions of the pedestrians. Let K be the number of pedestrians in an area, and let $\theta_1, \theta_2, \ldots, \theta_w$ ($w = K(K-1)/2$) be the difference in walking directions of each pair of the K pedestrians. Given that θ_s is the $\lceil 0.7w \rceil$-th largest element among $\theta_1, \theta_2, \ldots, \theta_w$, the subcategories are defined as:

- *High density with smooth flows*: The situations that satisfies $\theta_s < 45°$, where most of the pedestrians walk towards similar directions.
- *High density with intersections*: The situations where θ_s defined above is no less than $45°$.


(a) low density


(b) medium density


(c) high w/ smooth flows


(d) high w/ intersections

Figure 2. Categories of Crowd Density

Clearly, the areas with *High density with intersections* would arise higher stress and safety risks for the pedestrians, and should be avoided in the pedestrian navigation.

SYSTEM DESIGN

In order to identify the acceleration-based features that well reflect the characteristics of walking motion and ambient audio noise with different congestion categories, we have implemented a sensing application on the Android platform and conducted a preliminary experiment in an under ground shopping area near Umeda station, which is one of the most primary subway stations in Osaka area. The sensing application collects measurement data from accelerometers, magnetometers and gyro sensors every 20 milliseconds, and continuously records ambient audio with the built-in microphone. In this experiment, 12 student volunteers who hold an Android phone (Google Nexus S) in front of their body freely walked in a pedestrian flow. Through the experiment of totally 100 minutes, we collected the sensor measurements with each congestion category. Analyzing the sensor data from the preliminary experiment, we design two different congestion classification algorithms: the acceleration-based congestion classifier and the audio-based congestion classifier.

Acceleration-based Congestion Classifier

Acceleration-based Features

By analyzing the measurement data from accelerometers, magnetmeters and gyro sensors, we have found that time intervals between the walking steps has strong correlation with the congestion levels. Figure 4 shows typical examples of step intervals with different congestion categories. It can be seen that pedestrians walk with almost regular step intervals in the crowd with *low/medium density* or *high density with smooth flows*. On the other hand, under the congestion category of *high density with intersections*, the step invervals significantly vary since they often slow down or even stop to avoid collision with the surrounding people who walk toward a different direction. In addition, we can also see that the average step interval tends to be longer when the congestion category is *high density with smooth flows* and *high density with intersections*. This is because that the pedestrians need to move at similar speed with the surrounding people when walking through such a crowded area.

By analyzing the accelerometer readings from mobile devices, the walking steps can be robustly detected regardless of how the users hold their phones (*e.g.*, at hand, in a pocket and in a bag) [5]. Thus we have decided to employ the step intervals as acceleration-based features for congestion estimation.

Detecting Step Intervals

Step detection is a key component of our acceleration-based congestion estimation. Figure 5 shows vertical acceleration when a student volunteer walked in an open space, holding an Android phone (Google Nexus S) in front of his body. It can be seen that the original sensor readings contains small noise due to slight movements of the phone, which may incur misdetection of user's steps. To cope with the noise, we first apply a moving average filter with the window size of n, where average of the recent n acceleration samples is regarded as the current acceleration value. Unless otherwise noted, we assume $n = 10$ in the evalution section. The dotted curve in Figure 5 shows the vertical acceleration values after applying the moving average filter. The noise in the sensor readings is effectively mitigated and thus the rhythm of walking motion can be clearly observed.

It is known that the vertical acceleration takes a peak value at the moment when the foot contact the ground. Therefore, we detect the pedestrian's steps by finding the peak values in the vertical acceleration, and utilize the time intervals between each step as feature values for congestion estimation.

Acceleration-based Congestion Classification

Figure 6 shows the cululative distributions of step intervals with different congestion categories, which are calculated based on our preliminary experiment. With the *low/medium density*, more than 80% of the detected step intervals are less than 0.6 seconds. In contrast, the corresponding ratios with *high density with smooth flows* and *high density with intersections* are about 30%. Thus the step intervals tend to be larger when the crowd density is high. We can also see that the step intervals between 0.8 seconds and 3.0 seconds frequently occur under *high density with intersections*. As mentioned earlier, this is because the users often slow down or stop to avoid collision with the surrounding persons who walk toward a different direction.

Based on the observations above, we characterize the walking motion of the pedestrians by speeds and rhythms of the walking steps to classify the surrounding congestion of the clients into three categories. Whenever a client detects the user's step, it analyzes the step intervals within the window of recent N steps. If more than 80% of the step intervals in the window are less than 0.6 seconds, we regard that the user is walking at the *NORMAL* speed. Otherwise, the user

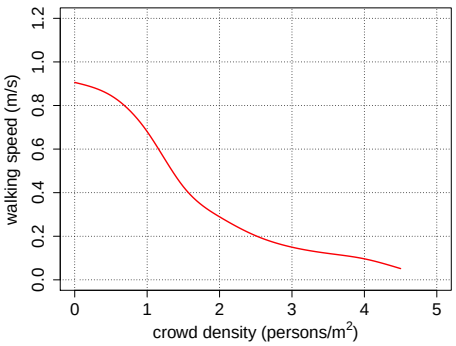


Figure 3. Relationship between crowd density and walking speed

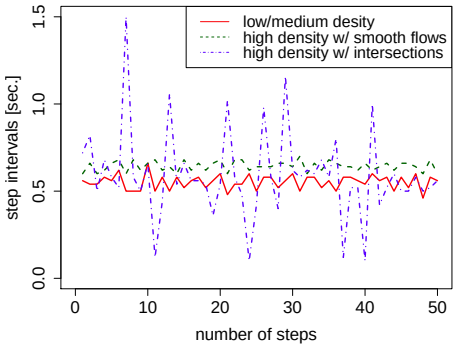


Figure 4. Examples of step intervals with different congestion categories

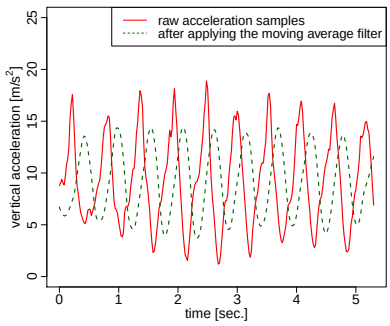


Figure 5. Vertical acceleration during walking motion

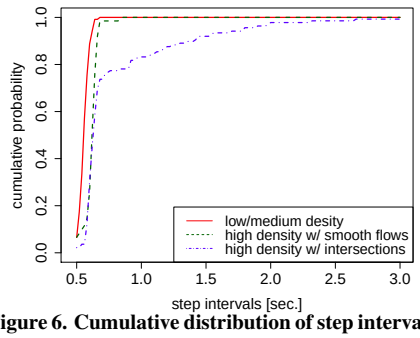


Figure 6. Cumulative distribution of step intervals

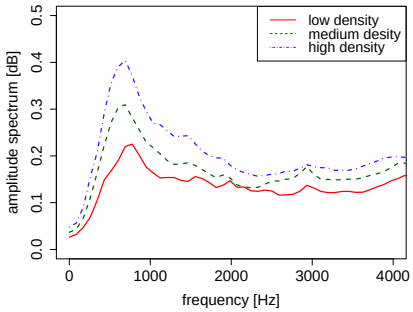


Figure 7. Frequency spectrum of ambient audio

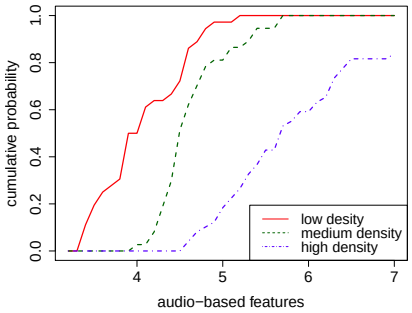


Figure 8. Cumulative distribution of audio-based features

Table 1. Rules for acceleration-based congestion estimation

speed	rhythm	congestion category
NORMAL	*REGULAR*	*low/medium density*
NORMAL	*IRREGULAR*	*low/medium density*
SLOW	*REGULAR*	*high density with smooth flows*
SLOW	*IRREGULAR*	*high density with intersections*

is regarded to be moving at the *SLOW* speed, where walking motion is constrained by the surrounding pedestrians. For the walking rhythm, if more than 20% of the step intervals are between 0.8 seconds and 3.0 seconds, we regard that the step rhythm of the user is *IRREGULAR*. Otherwise, it is defined to be *REGULAR*. We detect the speed and rhythm of the user's walking motion, and then estimate the current congestion category around the clients based on the rules in Table 1.

Audio-based Congestion Classifier

Audio-based Features

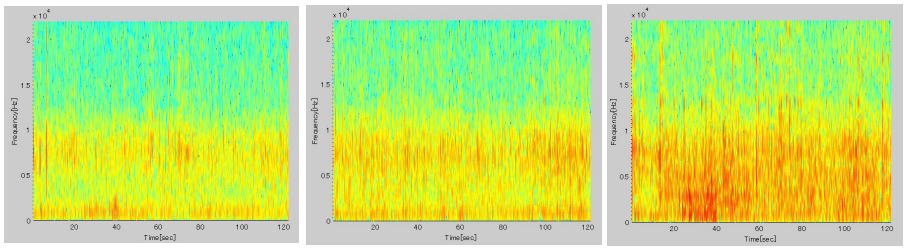
In order to find appropriate feature values for audio-based congestion estimation, we also analyzed the audio recordings that are collected in the preliminary experiment. Applying the fast fourier transformation (FFT) with a frame size of 20 milliseconds, we analyzed the power spectrum of the recorded audio samples. Figure 9 shows the spectrogram of the ambient audio under the three different congestion categories. The horizontal and vertical axes correspond to time and frequency, respectively, and colors at each point represent the amplitude of the frequency component (the red region represents the component with higher amplitude). As the crowd density increases, low frequency components below 10kHz exhibit larger power. To clarify this difference, we also show the average apmlitude of each frequency component (below 4kHz) over all the collected audio recordings in Figure 7. Due to the crowd noise, especially the frequency components below 2kHz get significantly larger, as the crowd density increases. Based on the observation, we extract the frequency components below 2kHz from the audio recordings of the recent 60 seconds, and the calculate the sum of all the amplitude values of these frequency components.

Audio-based Congestion Estimation

Using the feature value defined above, we classify the surrounding congestion of each client into *low density*, *medium density* and *high density*. Figure 8 shows cumulative distributions of the audio-based features with different congestion categories. We can see the distributions are clearly diffent depending on the surrounding crowd density, and thus can assume that the congestion levels could be reliably distinguished by applying some machine learing approach in the feature space. In this paper, we employ the k-nearest neighbor algorithm to construct a classifier to estimate the congestion categories based on the audio-based features. For the performance evaluation, we train the classifier with the feature values that are observed in our preliminary experiment.

Data Fusion at a Server

Each client i locally analyzes the acceleration readings and the audio recordings at their phone, and estimates the



(a) low density (b) medium density (c) high density

Figure 9. Spectrograms of ambient audio

Figure 10. Field experiment

acceleration-based congestion category $a_{i,t}$ and the audio-based congestion category $s_{i,t}$ of the surrouding region. Note that $a_{i,t}$ is updated at the timing of user's steps, while $s_{i,t}$ is obtained at every time slot of 60 seconds. Then the estimated congestion categories are periodically reported to a server with its current location $\boldsymbol{p}_{i,t}$. The reports from the clients are further analyzed by the server to classify the congestion at each area into four categories: *low density*, *medium density*, *high density with smooth flows* and *high density with intersections*. In order to improve the accuracy, the server applies the two-phase data fusion descibed below.

Data Fusion from Multiple Clients

At each time slot of T seconds, the server determines the acceleration-based congestion category and the audio-based congestion category of each area based on all the reports collected from the clients. Let $U_i = \{(\boldsymbol{p}_{i,t_1}, a_{i,t_1}), (\boldsymbol{p}_{i,t_2}, a_{i,t_2}) \ldots, (\boldsymbol{p}_{i,t_n}, a_{i,t_n})\}$ be the acceleration-based congestion categories that are observed by a client i during the recent T seconds. We also denote the set of congestion reports in U_i that are observed in the area A_k by $U_i^k \subseteq U_i$. For each area such that $U_i^k \neq \phi$, we find the acceleration-based congestion category, say $\tilde{a}_i^k$, that is the most frequently observed in U_i^k, and regard $\tilde{a}_{i,k}$ as the estimated category by the client i at this time slot. Collecting the estimates $\tilde{a}_{i,k}$ by all the clients i that have passed through the area A_k (*i.e.*, the clients such that $U_i^k \neq \phi$), we select the congestion category that are *voted* by the maximum number of clients as the acceleration-based congestion estimate at the current time slot. In the same way, we also determine the audio-based congestion category of each area based on the majority vote algorithm.

Fusing the Results from Different Sensors

Finally the server makes the current congestion estimate at each area by fusion of the acceleration-based congestion category and the audio-based congestion category that are estimated above. Our data fusion is based on the rules in Table 2. As we will mention in the discussion section, walking motion with each congestion category may be different among individual users; some users may frequently stop in uncrowded areas, while other users may pass through the crowd of people, keeping the constant walking speed. In contrast, the audio-based features would be less affected by such variance in the walking motion. Therefore, we basically put confidence in the acceleration-based congestion category in determining the final estimate. Exceptionally, we conclude that the congestion category is *unknown* when the acceleration-based congestion category is *low/medium density* and the audio-based congestion category is *high density*, since we can hardly classify the carrent congestion into the four congestion categories in such cases.

EVALUATION

Field Experiment

To examine the performance of our participatory crowd flow senising system, we have implemented the client application and conducted a field experiment in the underground shopping area near the Umeda subway station (in the same area as our preliminary experiment). In this experiment, 12 student volunteers walked along the pathway of 130m, holding Nexus S phones in front of their bodies (as in Fig. 10). The client application was run on the phones, and locally estimated the surrouding congestion category online. The estimated congestion category at each time was stored in the local strage of the phones, and we evaluated the effectiveness of data fusion at the server by analyzing these data after the experiment. By conducting such experiments 20 times, we evaluated the accuracy of congestion category classification at each client and at server side.

Congestion Estimation by Each Client

Acceleration-based Congestion Estimation

Table 3 shows the confusion matrix of the acceleration-based congestion estimation by each client. The rows and clumns in the table correspond to the actual congestion categories and the estimated congestion categories, respectively, and each cell in the table shows the ratio of the estimated congestion categories that the clients presented under each actual congestion category. Note that the diagonal elements in the table correspond to the accuracy rate of classification under each congestion category, while the sum of all the elements in each row must be 100%.

In all cases, the acceleration-based congestion classifier successfully identifies the congestion categories (*low/medium density*, *high density with smooth flows* or *high density with intersections*) around the clients with accuracy of more than 70%.

In the situations with *low/medium density*, the surrounding congestion of each client was misidentified as *high density with smooth flows* in 17.1% of the cases. This would be mainly due to a slope in the target area. The pathway that we conducted this experiment contains an upslope, where walking speed of the participants tended to slow down regardless

Table 2. Rules for data fusion

		Estimated category of the acceleration-based classifier		
		low / medium density	*high density w/ smooth flows*	*high density w/ intersections*
Estimated category of the audio-based classifier	*low density*	*low density*	*low density*	*low density*
	medium density	*medium density*	*medium density*	*medium density*
	high density	*unknown*	*high density w/ smooth flows*	*high density w/ intersections*

Table 3. Accuracy of acceleration-based local congestion estimation by individual clients

		estimated category		
		low/medium density	*high density w/ smooth flows*	*high density w/ intersections*
actual category	*low/medium density*	**74.7 %**	17.1 %	8.2 %
	high density w/ smooth flows	16.3 %	**75.2 %**	8.5 %
	high density w/ intersections	6.4 %	22.8 %	**70.8 %**

Table 4. Accuracy of audio-based congestion estimation

		estimated category		
		low density	*medium density*	*high density*
actual category	*low density*	**63.9 %**	33.3 %	2.8 %
	medium density	2.7 %	**78.4 %**	18.9 %
	high density	2.1 %	16.3 %	**81.6 %**

of the surrounding congestion levels. While their rhythms of walking steps were still regular, the system sometimes recognizes the slight increase of average step intervals to be caused by the surrouding congestion.

When the congestion category around the client is *high density with smooth flows*, the distinction from *low/medium density* failed with probability of 16.3%. This would be significantly due to the difference in walking motion among the participants. Although most of them slowed down their walking speed according the the speed of surrouding pedestrian flow, some participants passed through the surrounding crowd, keeping their original walking speed. Consequently, a few clients continuously presented the wrong congestion category, depressing the average accuracy.

For the situations with *high density with intersections*, the most confusing category was the *high density with smooth flows*. Our acceleration-based congestion classifier distinguishes these two categories by capturing the characteristic walking motion in avoiding conflict with the surrounding pedestians. Such motion would be usually, but may not be always, observed at crowded intersections. If the user smoothly passes through the intersection area, the intersection may not be detected by the clients.

All of the errors described above can be effectively mitigated by data fusion at the server, as we will show in the following sections.

Audio-based Congestion Estimation

Table 4 shows the confusion matrix of the audio-based congestion classifier. Except for the situations with *low density*, it could achieve the high estimation accuracy around 80%. The errors in *low density* would be significantly due to the surrounding audio noise that are not caused by the crowds (*e.g.*, the background music in the shopping area). Azizyan et al. [1] show that the characteritics of ambient audio are substantially different among the types of locations (*e.g.*, bookstore, boutique and pub). While we trained the audio-based congestion classifier by aggregating all the audio samples regardless of the locations where they recorded, the accuracy of the audio-based classifier would be further improved by constructing tailored classification models for each region. However, it incurs heavy training effort since we need to collect audio recordings for all the pairs of areas and congestion categories to construct such area-specific audio-based congestion classifiers.

Data Fusion from Multiple Clients

The server determines the acceleration-based congestion category and the audio-based congestion category of each area based on all the reports collected from the clients. Table 5 shows the accuracy of acceleration-based congestion estimation after the data fusion at the server. While some clients may report wrong congestion categories, the server effectively reduces the impact of such errors on the final congestion estimate by the majority-vote algorithm, and improves the congestion classification accuracy by 1.6–7.1%.

Although we also apply such data fusion for the audio-based congestion classifier, we could not observe any accuracy improvements. Since the ambient audio that are recorded by the clients in the same area is highly similar with each other, they usually present the same estimated category.

Integrating Results from Two Congestion Classifiers

Finally the server makes final decision on the congestion estimate at each area by fusion of the acceleration-based congestion category and the audio-based congestion category. Table 6 shows the confusion matrix after the data fusion process. As seen, our system could classify the congestion at each area into four categories with accuracy of 60–78%.

While the audio-based congestion classifier can distinguish the crowd density with finer resolution (*i.e.*, into three categories), it cannot capture the behavior of pedestrian flows. In contrast, the acceleration-based classifier can recognize the behavior of pedestrian flows (*e.g.*, the characteristic walking motion at intersections of multiple pedestrian flows), while it cannot capture the crowd density unless the walking motion of the users are affected by the surrounding pedestrians. Our system effectively combines these two classifiers to recognize detailed behavior of the pedestrian flows in the environment.

CONCLUSION AND FUTURE WORK

Table 5. Accuracy of acceleration-based congestion estimation after data fusin at a server

		estimated category		
		low/medium density	*high density w/ smooth flows*	*high density w/ intersections*
actual category	*low/medium density*	**80.3 %**	12.6 %	7.1 %
	high density w/ smooth flows	13.1 %	**82.3 %**	4.6 %
	high density w/ intersections	6.9 %	20.7 %	**72.4 %**

Table 6. Accuracy of congestion estimation (merged)

		estimated category				
		low density	*medium density*	*high w/ smooth flows*	*high w/ intersections*	*unknown*
actual category	*low density*	**63.9 %**	33.3 %	0.4 %	0.2 %	2.2 %
	medium density	2.7 %	**78.4 %**	2.4 %	1.3 %	15.2 %
	high w/ smooth flows	2.1 %	16.3 %	**67.2 %**	3.8 %	10.6 %
	high w/ intersections	2.1 %	15.3 %	16.9 %	**60.1 %**	5.6 %

In this paper, we have proposed a system for estimating crowd density and smoothness of pedestrian flows by participatory sensing with mobile phones. Based on preliminary experiments, we have modeled the walking motion and ambient noise with verious levels of crowds. By analyzing accelerometer readings and audio recordings obtained from mobile phones in a crowd, our system classifies the current behavior of the crowd at each area into four categories. Through field experiments using Android smartphones, we have shown that our system could recognize the current situation in the environment with accuracy of 60%–78%.

When the number of clients in the area is extremely limited, we can hardly expect sufficient accuracy gain by fusion of estimates from multiple clients. To cope with the problem, we plan to utilize history of congestion estimates, which can be stored at the server. Probability distribution of the past congestion estimates at the same time of day and at the same location would effectively complement the lack of clients. In addition, we will extend our crowd flow sensing system to do with other types of locations (*e.g.*, open space, stairs and ticket gates) as well as pathways. Utilizing other types of sensors (*e.g.*, magnetometers, gyro sensors and proximity sensing with Bluetooth) is also an important direction of this project.

ACKNOWLEDGMENTS

This work was supported in part by the KDDI foundation and the CPS-IIP Project (FY2012 - FY2016) in the research promotion program for national level challenges by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

REFERENCES

1. Azizyan, M., Constandache, I., and Roy Choudhury, R. SurroundSense: Mobile phone localization via ambience fingerprinting. In *Proc. MobiCom '09* (2009), 261–272.
2. Chintalapudi, K. K., Iyer, A. P., and Padmanabhan, V. Indoor localization without the pain. In *Proc. MobiCom '10* (2010), 173–184.
3. Enzweiler, M., and Gavrila, D. M. Monocular pedestrian detection: Survey and experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence 31*, 12 (2009), 2179–2195.
4. Kannan, P. G., Venkatagiri, S. P., Chan, M. C., Ananda, A. L., and Peh, L.-S. Low cost crowd counting using audio tones. In *Proc. SenSys '12* (2012), 155–168.
5. Li, F., Zhao, C., Ding, G., Gong, J., Liu, C., and Zhao, F. A reliable and accurate indoor localization method using phone inertial sensors. In *Proc. UbiComp '12* (2012), 421–430.
6. Lu, H., Pan, W., Lane, N. D., Choudhury, T., and Campbell, A. T. SoundSense: Scalable Sound Sensing for People-Centric Applications on Mobile Phones. In *Proc. MobiSys '09* (2009), 165–178.
7. Nakata, T., and Takeuchi, J.-i. Mining traffic data from probe-car system for travel time prediction. In *Proc. KDD '04* (2004), 817–822.
8. NAVITIME Japan Co., Ltd. The NAVITIME navigation system. `http://corporate.navitime.co.jp/en/`.
9. Older, S. *Movement of Pedestrians on Footways in Shopping Streets*. Traffic engineering & control, 1968.
10. Rana, R. K., Chou, C. T., Kanhere, S. S., Bulusu, N., and Hu, W. Ear-phone: an End-to-End Participatory Urban Noise Mappingsystem. In *Proc. IPSN '10* (2010), 105–116.
11. Takayanagi, H., Sano, T., and Watanabe, H. A study on the pedestrian occupied territory in the crossing flow : The analysis with pedestrian territory model. *Journal of Architecture and Planning*, 549 (2001), 185–191.
12. Vehicle Information and Communication System Center. Vics: Vehichle information and communication system. `http://www.vics.or.jp/english/vics/index.html`.
13. Wang, H., Sen, S., Elgohary, A., Farid, M., Youssef, M., and Choudhury, R. R. No need to war-drive: Unsupervised indoor localization. In *Proc. MobiSys '12* (2012), 197–210.
14. Weppner, J., and Lukowicz, P. Bluetooth based collaborative crowd density estimation with mobile phones. In *Proc. PerCom '13* (2013), 193–200.
15. Yin, J., Yang, Q., and Ni, L. M. Learning adaptive temporal radio maps for signal-strength-based location stimation. *IEEE Transactions on Mobile Computing 7*, 7 (2008), 869–883.