A Framework for Evacuation Hotspot Detection after Large Scale Disasters using Location Data from Smartphones: Case Study of Kumamoto Earthquake

Takahiro Yabe
University of Tokyo[1]
yabe0505@iis.u-tokyo.ac.jp

Kota Tsubouchi
Yahoo Japan Corporation[2]
ktsubouc@yahoo-corp.co.jp

Akihito Sudo
University of Tokyo[1]
sudoa@iis.u-tokyo.ac.jp

Yoshihide Sekimoto
University of Tokyo[1]
sekimoto@iis.u-tokyo.ac.jp

ABSTRACT
Large scale disasters cause severe social disorder and trigger mass evacuation activities. Managing the evacuation shelters efficiently is crucial for disaster management. Kumamoto prefecture, Japan, was hit by an enormous (Magnitude 7.3) earthquake on 16th of April, 2016. As a result, more than 10,000 buildings were severely damaged and over 100,000 people had to evacuate from their homes. After the earthquake, it took the decision makers several days to grasp the locations where people were evacuating, which delayed of distribution of supply and rescue. This situation was made even more complex since some people evacuated to places that were not designated as evacuation shelters. Conventional methods for grasping evacuation hotspots require on-foot field surveys that take time and are difficult to execute right after the hazard in the confusion.
We propose a novel framework to efficiently estimate the evacuation hotspots after large disasters using location data collected from smartphones. To validate our framework and show the useful analysis using our output, we demonstrated the framework on the Kumamoto earthquake using GPS data of smartphones collected by Yahoo Japan. We verified that our estimation accuracy of evacuation hotspots were very high by checking the located facilities and also by comparing the population transition results with newspaper reports. Additionally, we demonstrated analysis using our framework outputs that would help decision makers, such as the population transition and function period of each hotspot. The efficiency of our framework is also validated by checking the processing time, showing that it could be utilized efficiently in disasters of any scale. Our framework provides useful output for decision makers that manage evacuation shelters after various kinds of large scale disasters.

CCS Concepts
• **Information systems → Geographic information systems**
• **Information systems → Data mining**

Keywords
Disaster Management, Human Mobility, Urban Computing, Evacuation Hotspot Detection, Location Data

SIGSPATIAL'16, October 31-November 03, 2016, Burlingame, CA, USA
 ISBN 978-1-4503-4589-7/16/10 $15.00
DOI: http://dx.doi.org/10.1145/2996913.2997014

1. INTRODUCTION
Large scale disasters such as Magnitude 7 (M7) earthquakes, tsunamis, and mega-typhoons damage urban infrastructure and cause severe social disorder. As a result, many people evacuate to safe places. Efficiently managing the evacuation activities is crucial for decision makers after disasters. However, during the Great East Japan Earthquake, efforts to grasp the evacuation movements of the victims did not function well [25]. This resulted in inefficient distribution of supplies and rescue, causing inconvenience for the victims. This issue was also raised after the Kumamoto earthquake (M7.3) that occurred at 9:26PM of 14th April 2016. More than 10,000 residential buildings had collapsed, and unfortunately 49 people were killed mainly due to building collapse [1]. As we can observe from Figure 1, the largest shock (M7.3) struck near the densely populated central areas of Kumamoto, which forced over 100,000 people to move away from their homes to evacuation shelters [1,20].
The mass evacuation activities of the victims caused serious issues. Similarly to the Great East Japan Earthquake, grasping the locations of all the evacuation hotspots was extremely difficult in the chaos and confusion after the earthquake. This was made even

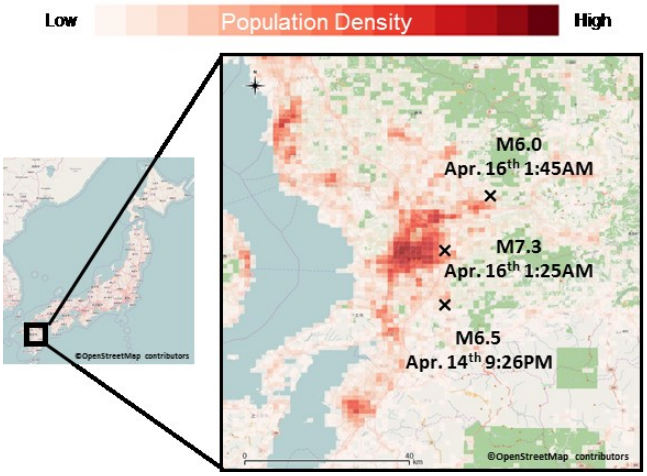

Figure 1. Crosses indicate epicenters of the three >M6 earthquakes. The thickness of red color in each grid cell represent the usual population density of each grid cell. The largest earthquake occurred near the highly populated area.

[1] 4-6-1 Komaba, Meguro-ku, Tokyo, Japan. (81)-3-5452-6009
[2] 9-7-1 Akasaka, Minato-ku, Tokyo, Japan. (81)-3-6864-3412

complex since some people evacuated to locations which were not officially designated as evacuation places, such as parking areas of large shopping malls and local parks. As a result, many evacuation hotspots which were not recognized by the administrative organizations as evacuation shelters couldn't be provided with food and supplies efficiently [2]. This increased the burden for the evacuees for several days.

Therefore, there is an urgent need for an efficient framework for estimating evacuation hotspots right after a natural disaster. The framework needs to require less time and less work load for the decision makers compared to the conventional on-foot field search for evacuation hotspots, since they are busy managing the situation after the earthquake.

Recently, GPS and call detail records (CDR) of mobile phones are being used for human mobility analysis [3,4,5,7], and are applied to various fields of study such as traffic management [11,12,14], urban planning [13], and pandemic simulations [16]. Some studies have analyzed the irregular human mobility after natural disasters such as Hurricane Sandy, Great East Japan Earthquake, and Haiti Earthquake [17,18,19], but none have proposed a framework for a real-time evacuation hotspot estimation.

In this paper, we propose a framework for estimating evacuation hotspots using location data from smartphones that works efficiently in disasters of any scale. To demonstrate the accuracy and efficiency of the framework, we used Yahoo! Japan's GPS dataset for detecting and analyzing the evacuation hotspots after the Kumamoto earthquake. Through the demonstration on the Kumamoto earthquake, we show that our framework accurately estimates evacuation hotspots, and also that this process can be completed at a significantly high speed and requires low effort compared to the conventional on-foot investigations. We then present some useful and insightful analysis that can be performed using the results obtained from our framework.

Our key contributions of this paper are as follows:

- We propose a framework to tackle hotspot detection problems following large scale natural disasters by using location data of smartphones.
- We validate the accuracy and efficiency of our framework by estimating the evacuation hotspots after the Kumamoto earthquake using actual GPS data provided by Yahoo Japan.
- We verify that outputs of the framework can be utilized by decision makers who manage evacuation shelters.

The rest of the paper is constructed as follows; in section 2, we explain the richness of the GPS dataset of Japan that we use in our framework, and also show that our dense GPS dataset can effectively estimate population distribution and capture sudden evacuation movement; in section 3, we explain our proposed framework; in section 4, we validate our framework by demonstrating in the Kumamoto earthquake; in section 5, we discuss our results; in section 6 we summarize the related works on mobility estimation and disaster response frameworks; and in section 7, we state our conclusion.

2. USED DATASET

To run this framework, we use two kinds of datasets. To estimate the people movement and evacuation hotspots, we use Yahoo! Japan's smartphone GPS dataset of around 1.2% of the population. As input parameters of our framework, we utilize the shapefile of Japan provided by the Geospatial Information Authority of Japan.

2.1 Mobile Phone GPS Dataset

The Yahoo Japan Disaster App collects the GPS data of smartphones of individuals who have agreed to provide their location data to Yahoo! Japan when installing the app. Each GPS record contains an anonymized user ID, longitude, latitude, and timestamp. The GPS data are collected every day when the individuals move around, while the smartphone is turned on. In total, GPS data of around 1 million individuals (sample rate around 1% from all over Japan) have been collected, which makes Yahoo Japan's GPS dataset one of the richest datasets in the world. As shown in Table 1, for the experiment, we use a total of 22,124 users' 418,119 total GPS logs from a period of January 1st to May 16th of 2016 which were located in Kumamoto area. The data is dense to the extent that plotting all the GPS points taken on a single day, as shown in Figure 2, can draw the entire shape and road network of the Kumamoto area. This figure shows the richness and density of Yahoo Japan's GPS dataset. We use this data to estimate the evacuation hotspots in our framework, and to contribute to a more efficient evacuation hotspot management system after large disasters.

Table 1. Number of unique IDs, number of logs of Yahoo Japan's GPS data taken from users' smartphones

Period of data	Average daily number of IDs in Kumamoto area	Average daily total GPS logs in Kumamoto area
2016/01/01 ~ 2016/05/16	22,124 (1.2% sample rate)	418,119 (avg. 19 logs/user/day)

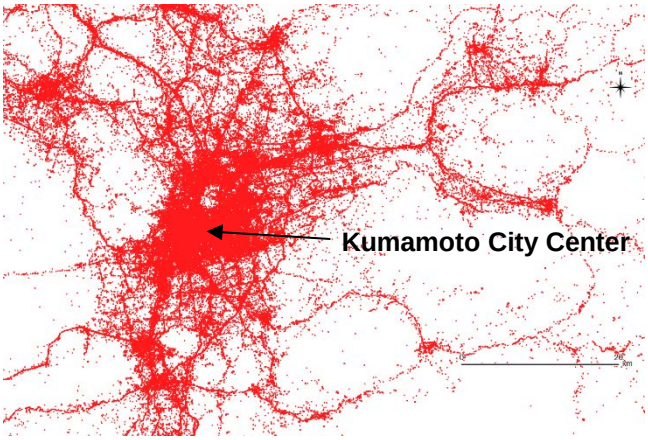

Figure 2. Collected GPS data points on April 1st 2016 of Kumamoto area. We can clearly observe the congested city center areas and also the road networks connecting the clusters, by plotting the dense GPS data of smartphones.

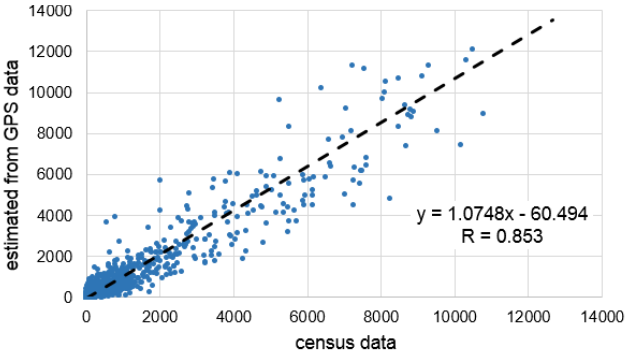

Figure 3. Comparison between census population data and population estimated from GPS data. A correlation of R=0.853 between them shows GPS dataset's goodness for population estimation.

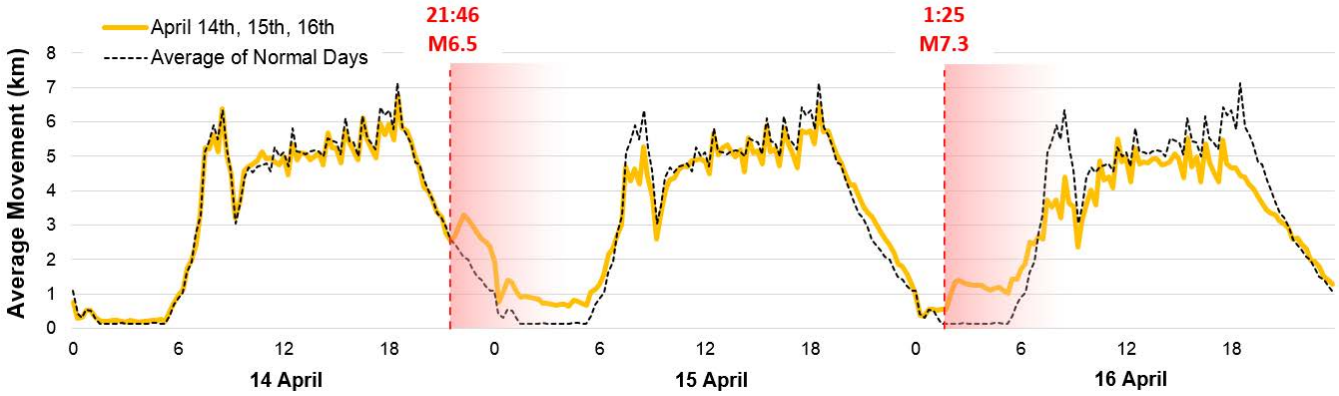

Figure 4. Average movement of individuals in Kumamoto area calculated from GPS trances of individuals. We can clearly observe a significant increase in movement after the 2 earthquakes compared to usual average movement, indicating the evacuation activities. The drop in movement after the 2nd earthquake shows the paralysis of the transportations systems.

2.1.1 Comparison of Census Data and Population Distribution Estimation from GPS Data

In this sub-subsection, we show that using GPS data, we can accurately estimate population distribution. Deville *et al.* [23] has shown in their research that nation-wide population distribution can be accurately estimated from mobile phone data, and we prove this for Yahoo! Japan's GPS data as well. Since GPS data has a sample rate of around 1%, we magnify each point so that the total number of points equals the actual value. We divide Kumamoto area into 1km grid cells and aggregate the population in each grid cell. Figure 3 shows the scatter plot for actual (census-based) and estimated (from GPS data) population. Each point shows the population for a single grid cell. The correlation coefficient is 0.853, showing that GPS data of around 1% represents the population distribution very well, despite the 1% sample rate and the biased sampling rate between age groups (younger generations use smartphones and apps more than elder generations).

2.1.2 Average Movement Analysis

In this sub-subsection, we show that GPS data could also effectively capture sudden evacuation movement of the victims, despite the time length between the observations (average of around 1.5 points per hour). Using the GPS traces of each individual, we calculate the average movement (km/h). Figure 4 shows the average movement of individuals in Kumamoto area between April 14th April and 16th 2016, compared to average movement of normal days before the earthquakes. It is clarified that after the 2 large earthquakes (shown in red dotted lines), a sudden increase in movement of the people was triggered compared to usual days. We can also observe a decrease in movement after the disaster on the 16th, due to the collapse of transportation systems and lack of usual movement such as going to schools and offices.

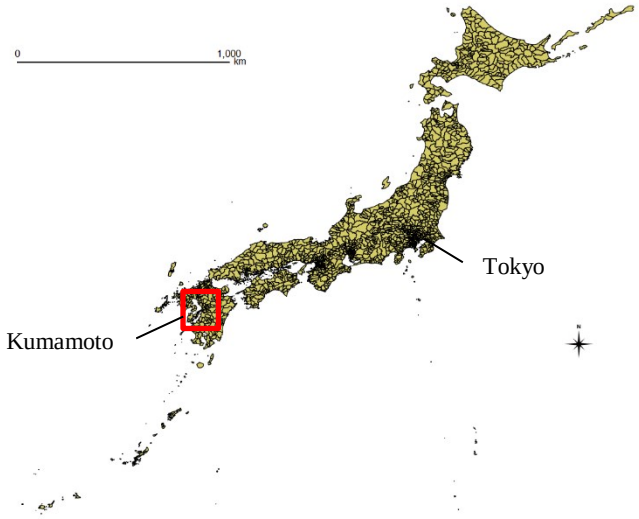

Figure 5. Shapefile of Japan

2.2 Shapefile of Cities in Japan

To extract the individuals who are staying in the disaster affected area from the GPS dataset, we use a shapefile of the affected area. A shapefile of Japan and its municipal governments is freely distributed at the website of Geospatial Information Authority[3] (GIA) of Japan. It includes not only the geographical shape of the municipal territories but also the census population information. Figure 5 shows the shapefile of Japan and Kumamoto area in red square.

3. PROPOSED EVACUATION HOTSPOT DETECTION FRAMEWORK

As shown in Figure 6, our proposed framework is consisted of the manual parameter input process and the automated calculation process. The manual parameter input could be completed momentarily, since shapefiles are available online, and the parameters are easy to input. The automated calculation process is consisted of 4 parts; 1) location data collection, 2) aggregation and smoothing the collected GPS data, 3) calculating the anomaly value of each grid cell, and 4) visualization of estimated evacuation hotspots. This framework is performed iteratively after the occurrence of the disaster, to update the output locations of evacuation hotspots. Our framework is efficiently designed so that it could be quickly processed, and also has little burden for the framework users since most of the parts are automated.

3.1 Manual Input Parameters

Our system requires three input parameters to function; a shapefile of the disaster hit location, the grid cell size of the visualization map, and the time and dates for results output. These three input parameters can be easily obtained and entered, making the system very easy and efficient to function even in the confusion after disasters.

3.1.1 Shapefile of Disaster Hit Location

To calculate the evacuation hotspots caused by the disaster, the user needs to input the shapefile of the location where the disaster

[3] HTTP://WWW.GSI.GO.JP/ENGLISH/INDEX.HTML

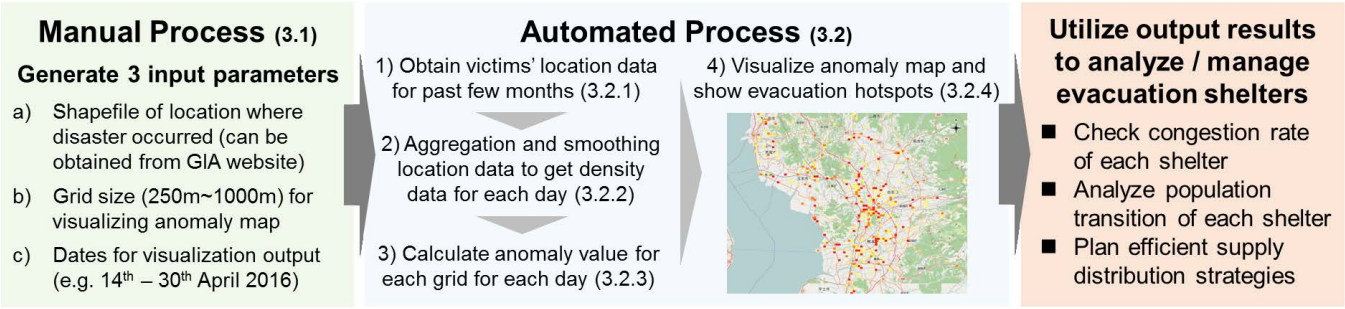

Figure 6. Diagram of our proposed hotspot detection framework. After the occurrence of the disaster, we manually input the a) shapefile of location where disaster occurred, b) the preferred grid cell size for visualization, and c) the dates for the visualization output. All the following processes are automated. The system calculates the anomaly value for each grid cell and visualizes the evacuation hotspots. This output contributes to a more efficient supply and rescue distribution plan. The numbers in brackets are section numbers explaining that part of the framework.

occurred. Shapefiles are needed to extract the GPS data of the individuals who are inside the disaster affected area. Shapefiles of Japan, shown in Figure 5, can be obtained freely from GIA website.

3.1.2 Grid cell Size for Visualizing Maps

Users need to input the grid cell size for outputting the anomaly maps. The grid cell sizes can be set as either 100m, 250m, 500, or 1000m. This parameter needs to be selected according to the wideness of the disaster affected area. For example, in Kumamoto earthquake, the total affected area was around 60km x 40km, therefore the grid cell size was set as 1000m. The grid cell size affects the processing time of the framework. Therefore, it is needed to check the size of the affected area and select the appropriate grid cell size. Analysis of processing time is shown in section 4.4.

3.1.3 Time and Dates for Visualizing Maps

The last parameters are the time and dates for outputting the anomaly maps. The default time is set for midnight to observe where people are staying the night, but can be changed to a different time if needed. Also, the dates for visualizing the maps are required. It is recommended to output the maps of days before the disaster occurrence day as well as after the disaster, to compare the rate of irregularity of congestion of before and after the disaster.

3.2 Automated Data Processing

Our framework is aimed to estimate evacuation hotspots right after the disaster, meaning that it has to function efficiently even in chaos situations after natural disasters. Therefore, we automated a large part of our framework. The blue part in Figure 6 shows the automated part, which is consisted of 4 processes.

3.2.1 Location Data Collection

Location data such as GPS data of the Yahoo! Japan app users are collected continuously, and they are stored in an internal server. Once a disaster occurs, the GPS logs recorded within a period of a few months before the disaster are collected. Then, logs located near the disaster hit area are extracted for each day. For spatial extraction, shapefile data of the area of disaster occurrence are needed as input data. This data collection phase can be completed in a few hours.

3.2.2 Aggregation and Smoothing of GPS Data

We aggregate the collected GPS data to obtain the population distribution of each day. To calculate the population distribution at night, we aggregate the last log for each ID (where he/she is staying the night) into grid cells of 100m~1000m size to anonymize and maintain the privacy of the users. Then, to overcome the relatively low sample rate (around 1%), we perform a Gaussian kernel density estimation [22] given by equation (1, 2).

$$\hat{f}_h(x,h) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right) \quad (1)$$

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}x^2} \quad (2)$$

Where there are n grid cells in the disaster hit area, x is the original population value, h is the bandwidth, and $f_h(x)$ is the estimated population value. We aggregate and smooth the population distribution data for all the days where data are available. The parameter *h*, called the bandwidth, is empirically decided, in the following sub-subsection.

3.2.3 Calculation of Bandwidth for Gaussian Kernel Density Estimation

To smooth the aggregated population density, we use the Gaussian Kernel Density Estimation method. The method requires the bandwidth parameter *h*, which decides the wideness of area when smoothing the values. If we set the bandwidth too large, the values of the grid cells would be strongly smoothed, ultimately setting the same value for all grid cells. On the other hand, if we set the bandwidth too small, none of the values would be smoothed, resulting in the same results as the original set of values. We empirically determine the value of h by finding the value that minimizes the error between GPS based usual population data of grid cell i ($\widehat{f_{h,i}}(h)$) and census population data of grid cell i ($f_{c,i}$).

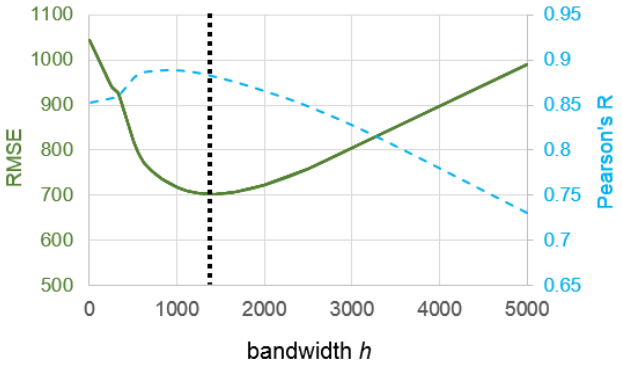

Figure 7. Bandwidth *h* and the RMSE/Pearson's R of the Gaussian Kernel Density Estimation results. h=1400m had the least error of RMSE = 704, compared to the original results (RMSE = 1085).

We have to note that the census data is not completely accurate, however can be agreed as the dataset with little error. We calculate the root mean square error (shown in equation (4)) and Pearson's R (shows in equation (5)), using the average estimated population $\overline{\widehat{f_h}(h)}$ and average census population $\overline{f_c}$, to evaluate the bandwidth h. Results are shown in Figure 7, where h=1400m showed the lowest RMSE (=702 people) among all values. Smoothing the grid cell population data can reduce the error between GPS aggregated data and the official census data, and increase the accuracy of the population estimation compared to the results without smoothing (RMSE = 1085, R = 0.853).

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\widehat{f_{h,i}}(h) - f_{c,i})^2} \quad (4)$$

$$\text{R} = \frac{\sum_{i=1}^{n}(\widehat{f_{h,i}}(h) - \overline{\widehat{f_h}(h)})(f_{c,i} - \overline{f_c})}{\sqrt{((\sum_{i=1}^{n}(\widehat{f_{h,i}}(h) - \overline{\widehat{f_h}(h)})^2)(\sum_{i=1}^{n}(f_{c,i} - \overline{f_c})^2))}} \quad (5)$$

3.2.4 Calculation of Anomaly Value

The anomaly value $K_{i,j}$ of each grid cell i on day j in the Kumamoto area are calculated by equation (3), using the average population μ_i of grid cell i, the standard deviation σ_i for each grid cell i, and the population in grid cell i on day j, denoted by $M_{i,j}$. This anomaly value measures the relative congestion of the grid cell compared to an average day, by considering the fluctuation of the population in that grid cell. Therefore, the anomaly value $K_{i,j}$ is dependent on the divergence of the grid cell's usual population. Using this metric, we can compare the anomaly in all of the grid cells fairly, despite the usual population difference between the grid cells.

$$K_{i,j} = \frac{M_{i,j} - \mu_j}{\sigma_j} \quad (3)$$

3.2.5 Visualization of Evacuation Hotspots

After calculating the anomaly value of the grid cells in the disaster-hit area, we visualize the anomaly values onto a map, so that decision makers and administration organizations can check the evacuation hotspots quickly and easily. Visualization will be performed on a free GIS software called QGIS[4], and OpenStreetMap[5] would be used as the background map. The csv file containing the population data and anomaly value of each grid cell is generated, so that policy makers can execute further analysis on the evacuation hotspots. We will demonstrate some further analysis in section 4.3.

4. CASE STUDY OF KUMAMOTO EARTHQUAKE

To demonstrate and to test our framework, we estimate the evacuation hotspots with real time GPS dataset after the Kumamoto earthquake. First we empirically calculate the best bandwidth for Gaussian Kernel Density Estimation. Then, using that bandwidth, we visualize the anomaly maps, and validate our estimation by looking at the features of each of the hotspots. In addition, we perform additional analysis using the output results to clarify the population transition of each evacuation hotspot and their function length periods, which will be useful for decision makers who manage evacuation shelters after disasters. Finally, we verify the efficiency of our system by calculating the time taken for estimation.

4.1 Visualization and Analysis of Anomaly Value

Figure 8A shows the map of Kumamoto area with grid cells with $K_{i,j}$>3 colored in red, 2<$K_{i,j}$<3 colored in orange, and 1<$K_{i,j}$<2 in yellow. We can observe very few grid cells with $K_{i,j}$>3 on April 1st before the earthquake, meaning that the majority of the grid cells have a population within usual range. However on the 18th, after the large earthquake, there is a significant increase in the number of grid cells with high anomaly values, especially near the city center and the southern part where many people evacuated. These grid cells indicate the "evacuation hotspots", where people evacuated at a significant rate compared to the usual population in that grid cell. It is also interesting to observe $K_{i,j}$>3 grid cells located on roads near the coastline. The high anomaly values in these grid cells infer that many people stayed in their cars away from their houses to avoid being injured by building collapses. However, after a month from the earthquake, we can see a decrease in congested areas in Kumamoto area. This implies that many evacuees returned home (if their house was not completely damaged) or moved away to other areas of Japan for shelter, and many outsiders who were gathering in Kumamoto to volunteer work went back to their homes. Using these anomaly maps, we can detect locations that are

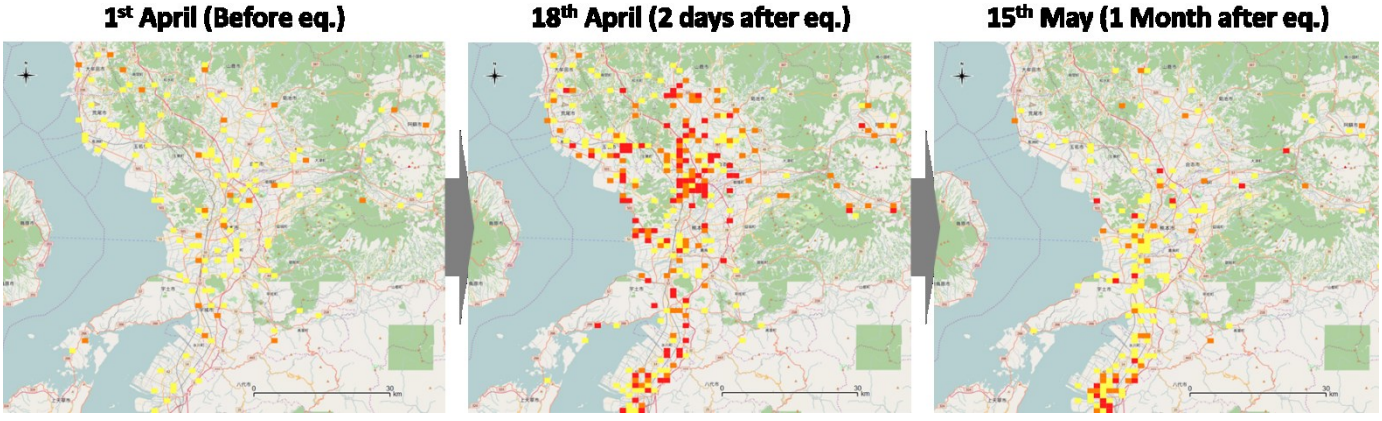

Figure 8A. Map of Kumamoto area; before, 2 days after, and 1 month after the earthquake. The grid cells with anomal values K>3, 2<K<3, and 1<K<2 are colored red, orange, and yellow, respectively. We can observe the increase in irregularly congested grid cells right after the earthquake, and its decrease as time passes and start to get back to normal from the shock.

[4] HTTP://WWW.QGIS.ORG/EN/SITE/

[5] HTTPS://WWW.OPENSTREETMAP.ORG/

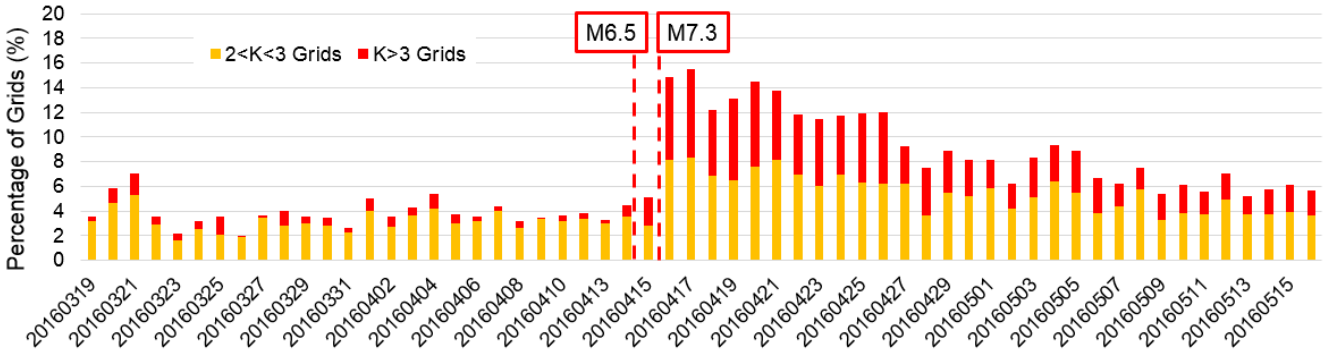

Figure 8B. Percentage of K>3, 2<K<3 anomaly grid cells by days. In normal days before the earthquakes, around 1% of K>3 and 3% of 2<K<3 grid cells are observed. However, after the shocks, irregularly congested grid cells increase at a significant rate. As time passes, the number of anomaly grid cells decrease, and by May 15th, the percentage of anomaly grid cells almost reach normal state.

significantly congested with people in real time, which can be utilized for rescue and supply distribution strategies. Using our framework, these maps will be easily obtained by the decision makers, by just entering three parameters explained in 3.1.

Figure 8B shows the temporal transition of the percentage of $K_{i,j}$>3 and 2<$K_{i,j}$<3 grid cells. Analysis of anomaly value in usual days show that the probability distribution of usual anomaly values follow a normal distribution. Therefore, before the earthquake, around 0.15% of the grid cells have $K_{i,j}$>3, and around 2.5% of the grid cells have 2<$K_{i,j}$<3. However after the earthquake, more than 7% of the grid cells had a $K_{i,j}$>3, and 8% had 2<$K_{i,j}$<3, resulting in a significant (= around 35 times larger probability) increase in high density evacuation hotspots compared to usual days. We can clearly see the impact of the earthquakes through observing the anomaly values of the grid cells in Kumamoto. Also, we are able to observe gradual return to normal state as time passes. By May 15th, around 2% of the grid cells have anomaly values over 3. Although this percentage is still significantly larger than normal states (0.15%), it shows that the situation in Kumamoto have recovered to a certain extent where many of the people were able to either return to their homes or evacuate to more secure places such as their relatives' houses.

Using our hotspot estimation method, not only were we able to estimate the evacuation hotspots of each day after the earthquake, but also grasp the impact of the earthquake and the length of time it takes for the damaged city to recover from the disaster.

4.1.1 Features of Evacuation Hotspots

In this sub-subsection, we validate our estimation of evacuation hotspots by checking the facilities located in each grid cell using Google map[6]. Figure 9 shows a map of central Kumamoto area and the grid cells with $K_{i,j}$>3 colored in red. The types of features located in each of the grid cells are written beside the grid cell. Many of the estimated hotspots contain facilities that have the capacity to contain large evacuation population, such as schools, city halls, convention centers and parks. From these results, we can conclude that our framework can accurately estimate the evacuation hotspots after the earthquake, and that our framework provides a convenient interface that shows the decision makers what features are congested with evacuees. We also have to note that out of the 1100 hotspots, around 100 were not designated as evacuation shelters before by the prefectural or the municipal government. This is an issue that made the detection of evacuation shelters difficult for the decision makers after the earthquake. As a result, people who evacuated to these non-designated facilities

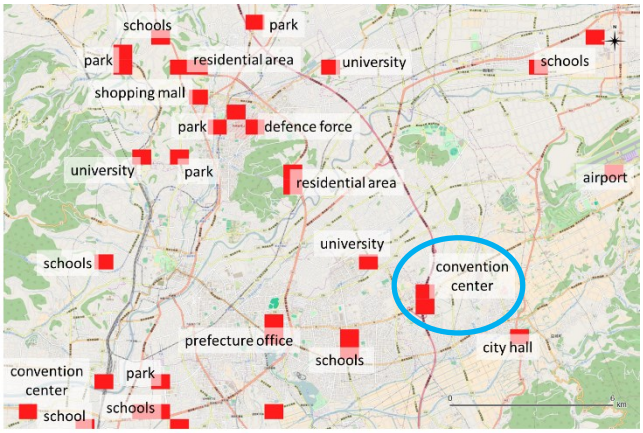

Figure 9. Map of central Kumamoto with K>3 grid cells colored in red. Close investigation on these grid cells revealed that most of these grid cells contain features that have a large capacity for evacuees such as schools, city halls, and convention centers. The blue circle shows Grand Messe Kumamoto (GMK).

could not be satisfied with much supplies compared to people who evacuated to designated shelters. To prevent these issues in the future, all the features highlighted in Figure 9 need to be designated as evacuation shelters, in all the areas that are prone to earthquakes in the world.

To further verify our results, in the next subsection, we will focus on two areas that showed a high rate of congestion ($K_{i,j}$>3) after the earthquake. The two areas are: an area that includes a convention center called Grand Messe Kumamoto, which was not designated as an evacuation shelter, and Mugishima area that includes two schools located south of the central Kumamoto area, which was designated as evacuation shelters before the earthquake. They both provided places to stay to a large number of evacuees, however showed different population transition patterns due to their status of designation.

4.2 Population Transition of Hotspots

To efficiently manage evacuation shelters after large disasters, not only grasping the evacuation hotspots but also the population transition in each evacuation hotspot is crucial. In this subsection, we take a closer look at some of the evacuation hotspots to verify

[6] HTTPS://MAPS.GOOGLE.CO.JP/

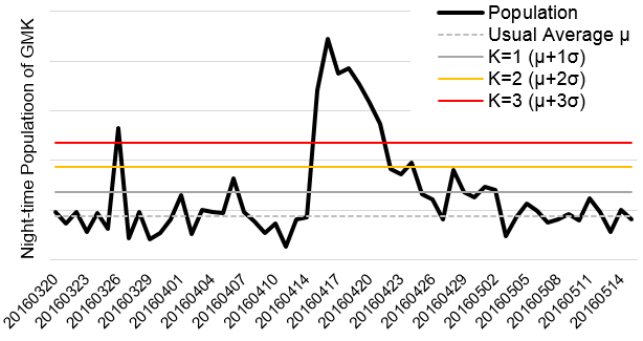

Figure 10A. Transition of population density near "Grand Messe Kumamoto", a large shopping mall which was not designated as an evacuation place. We can observe high congestion of K>3 after the earthquake, on April 15th ~ 21st.

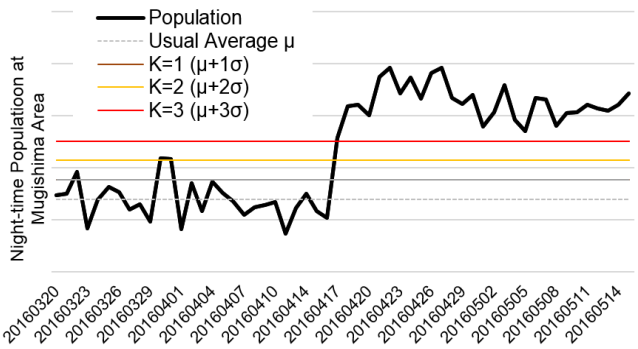

Figure 10B. Transition of population density near "Mugishima Elementary School", which was designated as an evacuation shelter. We can observe high congestion of K>3 even 1 month after the earthquake, at May 15th.

the accuracy of our estimation of the transition of population in the hotspot areas.

4.2.1 Analysis of Grand Messe Kumamoto Area

To analyze the evacuation activities with more detail, we focus on an evacuation hotspot where $K_{i,j}$>3 anomaly was detected, and plot the transition of the daily population in that selected hotspot. We focus on an area that includes "Grand Messe Kumamoto (GMK)" (circled in blue in Figure 9), a convention center in Kumamoto area, and analyze its daily transition of population near that facility. According to newspaper articles [20], many evacuees gathered in the parking area of GMK right after the earthquake, despite the fact that this facility was not designated as an evacuation location. The broad black line in Figure 10A shows the daily transition of the night-time population in GMK from March 20th to 14th May 2016. The dotted gray line shows the average population μ of usual days near GMK, and the gray, orange, and red lines indicate the $K_{i,j}$=1, $K_{i,j}$=2, $K_{i,j}$=3 lines for GMK, respectively. We can observe a rapid increase after the earthquakes on April 16th, and a significant anomaly in population density for more than a week over $K_{i,j}$=3. We can also spot an instantaneous $K_{i,j}$>3 on Saturday, March 26th. There was actually a large music festival held on this day at GMK, which is an example of an anomaly on a "usual" day. After April 18th, we can observe a quick decrease of population in GMK, and a gradual return to a usual level of population. By the beginning of May, the population in GMK has transferred back to the normal state. The evacuation hotspot in GMK was diminished after just a week after the earthquakes.

The increase of population on April 16th coincides with the information on the newspaper article [20]. We can conclude that the population in GMK, one of the evacuation hotspots, was accurately estimated from GPS dataset, and that the function as an evacuation shelter was finished after only one week from the earthquake.

4.2.2 Analysis of Mugishima Area

One of the evacuation shelters that attracted many evacuees was Mugishima elementary school in Mugishima area, which is located in the south of Kumamoto city. The broad black line in Figure 10B shows the daily transition of the night-time population in Mugishima area from March 20th to 14th May 2016. The dotted gray line shows the average population μ of usual days, and the gray, orange, and red lines indicate the $K_{i,j}$=1, $K_{i,j}$=2, $K_{i,j}$=3 lines, respectively. In this area, $K_{i,j}$>3 anomaly was detected from the April 18th, and has continuously been congested with people until 15th of May, which is 1 month after the earthquake. This result correlates with the fact that Mugishima elementary school was designated as one of the largest evacuation centers by Mugishima city before the earthquake with a capacity of more than 1500 people, and also that Mugishima East Park also hosted many evacuees after the shocks. We have to note that, unlike Grand Messe Kumamoto, shelters in Mugishima area continued to host many evacuees even after a month from the earthquake, and are still functioning as an evacuation shelter after a month. In the next sub-section, using the results obtained from our framework, we calculate the functioning period of all of the evacuation shelters. To better understand the functionality of the shelters, we analyze them by checking the geospatial location and designation status of the shelters.

4.3 Functioning Time Length of Hotspots

Through the case studies in 4.2.1 and 4.2.2, we have shown that the length of the functioning period differs between shelters. In this subsection, using the results obtained from our framework, we analyze the functioning period of the evacuation shelters from two perspectives; geospatial location and designation status. The analysis of each evacuation shelter's functioning length period gives useful information to decision makers who need to manage evacuation shelters after large disasters.

4.3.1 Geospatial Location and Function Period

To efficiently manage the shelters, there is a need to understand the functioning period of all the shelters. Figure 11A shows a map of Kumamoto area, showing the days of function of each shelter area.

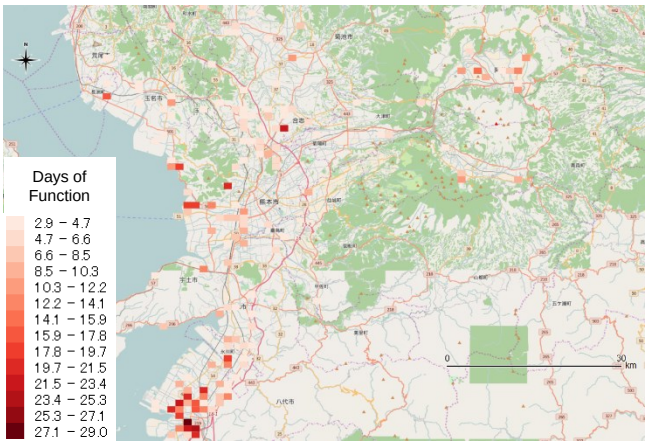

Figure 11A. Days of function of all evacuation shelters by geospatial location. Evacuation areas in the outskirts function for a longer period than shelters in the central parts of Kumamoto.

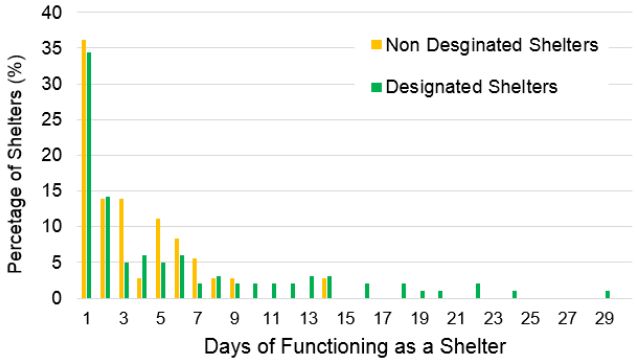

Figure 11B. Histogram of the days of function of evacuation shelters by designation of the government.

As we can observe, most of the shelters in the southern area, including Mugishima area, functioned for a long period (of around 20 days or more) compared to other areas. By comparing this map with Figure 1, we can infer that shelters in highly populated areas such as the city center of Kumamoto did not function for a long time, compared to shelters on the coastline and in the eastern mountain parts. From this result, we can conclude that shelters in the less populated areas, which host relatively poorer evacuees with weak structured houses, need to function for a longer period than in highly populated areas where people tend to be richer and have the ability to find their own places to stay, or fix their houses quickly after being damaged by the earthquake.

4.3.2 Designation and Function Period

As we have shown in the case studies, designated shelters and non-designated shelters both hosted many evacuees after the earthquake. However, while Grand Messe Kumamoto only functioned for a week, Mugishima area functioned as an evacuation shelter for over a month. Our following analysis shows that one of the reasons for this is the designation of the shelters. Designated shelters tend to have a better chance of receiving a satisfying amount of supply from the government, while non-designated shelters are hard to manage, resulting in less supply and support by the government. Figure 11B shows a histogram of the number of days that each shelter functioned, by designation status. Yellow bars and green bars show the histogram of non-designated and designated shelters, respectively. Designated shelters have a longer functioning period, with a maximum of 29 days and 29% of the shelters functioning for more than one week, while non-designated shelters have a maximum of 14 days, and only 13% of the shelters function for more than one week. From these results, we can clarify the importance of designating the evacuation shelters that attract many evacuees after disasters, to prevent people suffering from lack of supplies and support.

As we have demonstrated in sections 4.3, by using the results obtained from our framework, we can analyze the functionality of each evacuation shelter and utilize these information for the efficient allocation of evacuees and strategic designation of evacuation shelters prior to disasters.

4.4 Efficiency of our Framework

Finally, we validate the efficiency of our framework. Through sections 4.1 to 4.3, we clarified that our framework can accurately estimate evacuation hotspots, and that the output results could be used to execute useful analysis on evacuation shelters for decision makers and shelter managers. For this framework to function in an actual disaster case, the processing time needs to be less than a few hours in total. Furthermore, to apply this framework to various

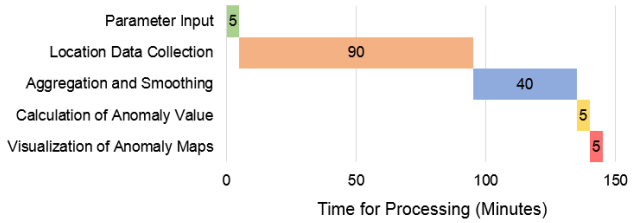

Figure 12A. Processing time of our framework. The time shown is for Kumamoto area (60km x 40km), and for 30 days including before and after the earthquakes.

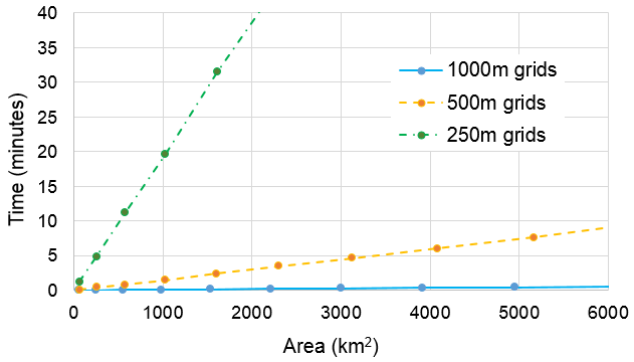

Figure 12B. Processing time of aggregating and smoothing the location data for 1 day. 3 lines show the time length for different grid cell sizes. While smoothing 1000m grid cells do not take much time even when the disaster affected area is large, 250m grid cells take a significant amount of time.

areas and types of disasters and provide the decision makers and shelter managers quickly, this framework needs to be robust for many various cases with different population size and area size. In this subsection, we validate the robustness of our framework by empirically showing the processing time by different area size.

Figure 12A shows the processing time for each step in our framework. For the processing, we used Intel Xeon CPU E5-2630L v3 1.80GHz, 128GB memory.

The 4 steps after parameter input are automated, therefore needs no work of the decision makers. The total processing time of the framework is around 145 minutes, and we can clearly see that the location data collection step and the data aggregation and smoothing step consumes the most time.

In the "location data collection" step, location data of individuals who are in the disaster affected area are obtained for around 30 days, including 29 usual days and the disaster occurrence day. In his step, a lot of the time is taken for geometrically searching IDs who are located in the disaster affected area, from the massive amount of location data (around 15.2 GB per day for entire Japan). The processing time for this step does not depend on the type of disaster or the wideness of the disaster affected area.

On the other hand, the processing time for "data aggregation and smoothing" step depends on the size of the disaster affected area and the size of the grid cells for aggregation. To apply our framework to various disasters that occur in any area of Japan, robustness against the size of the disaster affected area is required. Figure 12B shows the processing time for the "data aggregation and smoothing" step for different area sizes and grid cell sizes for 1 day. It is clarified that while smoothing 1000m grid cells does not consume time even when our framework is applied to a large area like Kumamoto earthquake, smoothing 250m grid cells consumes a significantly longer time to execute. We have to note that our

framework recommends to calculate the population density of around 30 days, meaning that the total consuming time would be 30 times longer than the value in Figure12B. Therefore it is not practical to aggregate and smooth into 250m grid cells in a large scale disaster. The Great East Japan earthquake and the following tsunami affected a significantly large area of Japan from the northern region to Tokyo, of around 1000 km^2, with 535 km^2 flooded. Using 1000m or 500m grid cell sizes, we can estimate the evacuation hotspots in the area hit even by an unprecedentedly large disaster using our framework.

5. DISCUSSION

Our framework for estimating evacuation hotspots using location data from smartphones provides useful information to decision makers and shelter managers quicker and with less effort than conventional methods. Providing useful and accurate information can contribute to making efficient supply distribution and rescue operation plans after disasters.

By calculating the anomaly value of each grid cell in Kumamoto after the earthquake, we were able to estimate the distribution of irregularly congested grid cells, despite the sample rate (around 1%). We were able to quantify the significant increase in number of grid cells with high anomaly values right after the occurrence of the earthquake, and statistically show the irregularity of the situation compared to usual days. In addition, the gradual decrease of grid cells with anomaly values higher than 2, showed the settling down of irregularity after a few weeks from the occurrence of the earthquakes. Through the analysis and visualization of the anomaly with our framework, we were able to observe the impact of earthquakes on the people's evacuation activities.

Secondly, by checking the features located in each grid cell that were estimated to be extremely congested, we verified the accuracy of our estimation. We were able to detect locations where administrative organizations had not designated as evacuation shelters, such as Grand Messe Kumamoto.

Through the analysis of evacuation shelters using the results obtained from our framework, we were able to analyze the geospatial distribution of the shelters that functioned for a long period, and also the relationship between designation by the government and the functionality of the shelters. This analysis can help decision makers to manage evacuation shelters and efficiently distribute the supplies, and to improve the lives of the people who have evacuated after the disaster.

To decrease the processing time and burden for the people who use the framework, we automated most of the processes and only parameter input has to be done manually. We validated the efficiency of our framework by calculating the process time with many different areas to cover, and with three different grid cell size types. We conclude that by choosing the right grid cell size, this framework can efficiently provide the results of evacuation hotspot estimation to decision makers after disasters that happen in any size or scale.

6. RELATED WORKS

6.1 Human Mobility Analysis using Mobile Phone Location Data

The increasing availability and fusion techniques of big and heterogeneous data has enabled the analysis for tackling major issues that cities face [8,9]. Due to the popularization of mobile phones, call detail record (CDR) and GPS data have become popular for analyzing people movement [3,4,5,7] and nation-wide population distribution [23]. Other studies have applied human mobility analysis to many fields such as traffic estimation [11,12,14], urban planning [13], and public health [16]. As for hotspot detection, Lukasczyk *et al.* [10] proposed a topological analytical approach, and Hong *et al.* [26] detected urban black holes using human mobility data.

6.2 Disaster Mobility Analysis

Recently, to understand the irregular people movement under disaster conditions, mobile phone CDR/GPS data are being used for analyzing calling activities after large disasters [6]. Human evacuation activities after natural disasters such as the Haiti earthquake [17], Hurricane Sandy [18], and the Great East Japan Earthquake [19] are also analyzed. However, these studies are commonly the analysis of people movement after a disaster and do not attempt to infer the real-time evacuation hotspots after the disaster.

6.3 Frameworks for Disaster Human Mobility Prediction

Chen *et al.* [27] proposed a framework using agent based simulation for predicting human mobility after disasters using CDR, however they used a simple potential model for the mobility prediction and did not attempt to estimate the evacuation hotspots. Similarly, Yabe *et al.* [24] proposed a system for predicting people movement in disasters using real time data. Horanont *et al.* [21] stated the potential usage of GPS data in emergency situations, but have not proposed an actual method or framework for evacuation hotspot estimation. We also have to note that the analysis was done after the settlement of the disaster in these two studies, and their frameworks' ability for prompt analysis and visualization right after the disaster have not been verified.

7. CONCLUSION

In this paper, we proposed a framework for estimating evacuation hotspots by calculating each grid cell's anomaly value after large disasters, using location data from smartphones. To the best of our knowledge, this framework is the first to focus on estimating evacuation hotspots, and the first to actually demonstrate the framework using real GPS dataset. Our framework can function quicker and with less effort compared to conventional methods that involve on-foot searches for evacuation centers where people are gathering.

To validate our method, we analyzed the population density anomaly after the Kumamoto earthquake (M7.3), and observed the sharp increase of high anomaly value grid cells in Kumamoto area caused by the evacuation activities of the victims. We then verified our estimation by looking at the features included in each grid cell, and also newspaper articles that mentioned the population transition in one of the evacuation hotspots. Further analysis on evacuation shelters using the estimated results have shown the usefulness of our framework's output for managing shelters.

Through the validation in the case study of Kumamoto and the efficiency test, we have confirmed the high accuracy and practicality of our evacuation hotspot framework using location data from smartphones.

8. REFERENCES

1. Japan Meteorological Agency, Evaluation of the Kumamoto Earthquake, (2016). (in Japanese)
2. West Japan Newspaper, Disparity between evacuation centers, April 20th (2016). (in Japanese)
3. Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A. L. Understanding individual human mobility patterns. Nature, 453(7196), (2008).
4. Sevtsuk, A., & Ratti, C. Does urban mobility have a daily routine? Learning from the aggregate data of mobile networks. Journal of Urban Technology, 17(1), 41-60. (2010).

5. Ashbrook, D., & Starner, T. Using GPS to learn significant locations and predict movement across multiple users. Personal and Ubiquitous Computing, 7, 5, 275-286. (2003).
6. Bagrow, J. P., Wang, D., & Barabasi, A. L. Collective response of human populations to large-scale emergencies. PloS one, 6(3) (2011).
7. Calabrese, F., Di Lorenzo, G., Liu, L., & Ratti, C. Estimating origin-destination flows using mobile phone location data. IEEE Pervasive Computing, 10(4), (2011).
8. Zheng, Y., Capra, L., Wolfson, O., & Yang, H. Urban computing: concepts, methodologies, and applications. ACM Transactions on Intelligent Systems and Technology (TIST), 5(3), 38. (2014).
9. Zheng, Y. Methodologies for cross-domain data fusion: an overview. Big Data, IEEE Transactions on, 1(1), 16-34. (2015).
10. Lukasczyk, J., Maciejewski, R., Garth, C., & Hagen, H. Understanding hotspots: a topological visual analytics approach. In Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems (p. 36). ACM. (2015)
11. Demissie, M. G., de Almeida Correia, G. H., & Bento, C. Intelligent road traffic status detection system through cellular networks handover information, Transportation research part C: emerging technologies, 32, 76-88. (2013).
12. Iqbal, M. S., Choudhury, C. F., Wang, P., & González, M. C. Development of origin–destination matrices using mobile phone call data. Transportation Research Part C: Emerging Technologies, 40, 63-74. (2014).
13. Wang, P., Hunter, T., Bayen, A. M., Schechtner, K. & González, M. C. Understanding road usage patterns in urban areas. Sci. Rep. 2, 1001 (2012).
14. Yang, Y., Gerstle, D., Widhalm, P., Bauer, D. & González, M. The potential of low-frequency data for the monitoring and control of bus performance. Transport. Res. Rec. J. Transport. Res. (2013).
15. Pentland, A. Society's nervous system: Building effective government, energy, and public health systems. IEEE Computer 45, 31–38 (2012).
16. Colizza, V., Barrat, A., Barthelemy, M., Valleron, A. J., & Vespignani, A.. Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. PLoS Med, 4(1) (2007).
17. Lu, X., Bengtsson, L., & Holme, P. Predictability of population displacement after the 2010 Haiti earthquake. Proc. National Academy of Sciences, 109(29), 11576-11581. (2012).
18. Wang, Q., & Taylor, J. E. Quantifying human mobility perturbation and resilience in Hurricane Sandy. PLoS one, 9(11), (2014).
19. Song, X., Zhang, Q., Sekimoto, Y., Horanont, T., Ueyama, S., & Shibasaki, R. Modeling and probabilistic reasoning of population evacuation during large-scale disaster. Proc. ACM SIGKDD (2013).
20. Kumamoto Daily Newspaper, Evacuation center at Grand Messe Kumamoto, April 24th (2016). (in Japanese)
21. Horanont, T., Witayangkurn, A., Sekimoto, Y., & Shibasaki, R. Large-scale auto-GPS analysis for discerning behavior change during crisis. IEEE Intelligent Systems, (4), 26-34. (2013).
22. Tukey, J. W. Exploratory data analysis. (1977).
23. Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F. R., Gaughan, A. E., & Tatem, A. J. (2014). Dynamic population mapping using mobile phone data. Proceedings of the National Academy of Sciences, 111(45), 15888-15893.
24. Yabe, T., Sudo, A., Kashiyama, T., Kanasugi, H., Sekimoto, Y. "Making Real-Time Predictions of People's Irregular Movement In a Metropolitan Scale under Disaster Situations", Proc. CUPUM (2015)
25. Cabinet Office of Japan, Issues on Emergency Rescue during Great East Japan Earthquake (in Japanese), (2012).
26. Hong, L., Zheng, Y., Yung, D., Shang, J., & Zou, L. Detecting urban black holes based on human mobility data. In Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems (p. 35). ACM. (2015)
27. Chen, F., Zhai, Z., & Madey, G. Dynamic adaptive disaster simulation: developing a predictive model of emergency behavior using cell phone and GIS data. In Proceedings of the 2011 Workshop on Agent-Directed Simulation (pp. 5-12). Society for Computer Simulation International. (2011).