ANALYSIS OF TEMPORAL CHANGES IN POPULATION DISTRIBUTION FOR DEVELOPING CITIES USING SOCIAL SURVEY DATA

Natsumi Ono [1], Yoshihide Sekimoto [2], Atsuto Watanabe [3], Teerayut Horanont [4], Ryosuke Shibasaki [5]

[1] Graduate Student, Graduate School of Frontier Sciences, University of Tokyo,
5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8568, Japan; Tel: +81-4-7136-4307;
E-mail: nono@csis.u-tokyo.ac.jp

[2] Associate Professor, Center for Spatial Information Science, University of Tokyo,
5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8568, Japan; Tel: +81-4-7136-4307;
E-mail: sekimoto@csis.u-tokyo.ac.jp

[3] Graduate Student, Graduate School of Frontier Sciences, University of Tokyo,
4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan; Tel: +81-3-5452-6417;
E-mail: atsuto@csis.u-tokyo.ac.jp

[4] Researcher, Department of Civil Engineering, University of Tokyo,
4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan; Tel: +81-3-5452-6417;
E-mail: teerayut@iis.u-tokyo.ac.jp

[5] Professor, Center for Spatial Information Science, University of Tokyo,
5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8568, Japan; Tel: +81-4-7136-4307;
E-mail: shiba@csis.u-tokyo.ac.jp

KEY WORDS: population distribution, developing countries, visualization, person trip data

ABSTRACT: Information about daily or hourly changes in population distribution is becoming increasingly important in business fields such as marketing and public services. In recent years, the prevalence of mobile phones has refined the data collection process; in certain countries, fragmentary data offered by mobile phone users are used to estimate and map the number of people in a certain area at any given time of a day. Such information would be potentially useful in cities of developing countries, where various infrastructure needs may not yet have been fully met. This study attempts to mimic the aforementioned process by using existing data from several such cities in order to examine how dynamic population distribution data could be used. Data are based on trip information from Person-Trip Surveys conducted by Japan International Cooperation Agency (JICA), and that have been reconstructed in order to identify the specific locations of people. Whereas past studies organized this information to represent the flow of people, this study attempts to clarify overall population distribution. Such visualization will help assist in the urban planning of cities, including establish what steps should be taken in the event of a natural disaster. Furthermore, implementing this method on various cities and comparing the results will help confirm what factors, cultural or otherwise, affect the temporal changes in urban population distribution. We hope this study will encourage further research, preferably with the collection of real-time location data from mobile phones.

1. INTRODUCTION

In recent years, information regarding the location of people is becoming increasingly available, and proving very useful in various situations. Not only is such information important in business fields such as marketing and public services, but it can also be valuable in the event of disasters, such as the earthquake and consequent tsunami that occurred in eastern Japan on March 11th of 2011. In order to have knowledge of the population distribution at any given moment, real-time location data should be collected, organized, and put through a visualization process.

1.1 Background

Previously published research includes various forms of population distribution data that cover areas around the world. For example, the Gridded Population of the World (GPW) was produced in 1995 by the National Center for Geographic Information Analysis (NCGIA) as the first raster global dataset of population (Pozzi et al., 2003). A second version (GPW2) was produced in 2000 by the Center for International Earth Science Information Network (CIESIN) at Columbia University and used a higher resolution. Data is gridded based on original census units, at the highest spatial resolution for which they are available by country, and uses two methods, the latter of which uses a smoothing method that assumes grids closer to units with high density are higher. Similar information is provided by LandScan, developed by the Oak Ridge National Laboratory (ORNL). LandScan models are tailored to match the data conditions and geographical natures of each individual country and region, and at approximately 1km resolution (30”×30”) it is currently the finest resolution global population distribution available. Whereas GPW provides night-time population, LandScan represents ambient population, meaning the average over 24 hours.

However, the recent and rapid spread of mobile phones is allowing us to obtain real-time information regarding the whereabouts of people. In Japan, there is an Internet site that continuously provides a “congestion map”, which uses fragmentary data offered by mobile phone users to estimate and map the number of people in a certain area at any given time of a day. A map of Japan is divided into units of approximately 300m, 600m, or 1.2km (depending on how closely the user zooms in on the online map), and each unit is shaded according to the population density at that moment. Furthermore, right-clicking a unit will provide a histogram showing the fluctuations in population of that unit over the past 24 hours. Such maps inform the general public about congestion levels in all areas, and may affect their choices in transportation and/or destination when going out, as well as be referred to for commercial purposes or for decision-making in the event of a disaster.

Maps representing real-time population distribution are not yet common in other countries, but such information would be potentially useful especially in cities of developing countries, where infrastructure needs have not yet been met and also where natural disasters may cause confusion. It would assist in the urban planning process, including prevention of secondary disasters.

1.2 Objectives

The ideal model would collect and organize real-time location data on a global scale. Difficulties may arise in technological application and in data collection, considering the differences in privacy and security policies of areas around the world. The objective of this study is to encourage research and eventual implementation by examining if dynamic population distribution data is indeed beneficial and if so, how it could be used. We attempt to imitate the aforementioned real-time “congestion map” in cities of developing countries, and analyze these maps to see how they can be used. In addition, creating maps for multiple cities and comparing them with one another would help clarify what factors, cultural or otherwise, may influence the temporal changes in urban population distribution. As mobile phone-based location data is currently unavailable, this study uses data from social surveys, but we hope the results of this study will encourage such data collection.

2. METHODOLOGY

For this study, we used data from person-trip surveys conducted by Japan International Cooperation Agency (JICA) and that have been reconstructed using an interpolation process by the People Flow Project at the University of Tokyo (which focuses on visualization of the actual flow of people as opposed to population density). Approximate locations, represented using longitude and latitude, are provided for every minute throughout the course of a 30 hour period. This study uses data from the three cities of Hanoi, Manila, and Jakarta, and parameters such as population, survey year, number of people and trips for each of these cities are listed in Table 1. In each survey between 1 to 6% of the population was sampled, with each person making 2 to 3 trips.

Next, we prepared grids consisting of 60 by 60 units, each of which was approximately 500 meters in length and width (15”×15”). Each grid map represented a certain minute in time, and thus using data spanning 30 hours produced a total of 1800 grids. The grids were centered on the central area of each city, and slight adjustments were made so that the most densely populated areas would be covered. Those data that did not fit these boundaries were discarded and excluded from the grid for that minute. The remaining valid data were allocated to their respective units, depending on their coordinates.

Finally, for the visualization process the completed grids were aligned with Open Street Maps (as the roads shown are the same ones used in the interpolation process). Units are colored according to the number of people inside of them, and for this paper five population levels were used, the lowest of which represents zero or no recorded people. In addition, histograms were produced for certain units in the grid maps.

Table 1. Basic information about person-trip survey data

City	Population (million)	Survey year	Number of people	Number of trips
Hanoi	7.16	2004	63,716	188,949
Manila	9.45	1996	231,889	471,035
Jakarta	2.10	2000	423,237	1,083,280

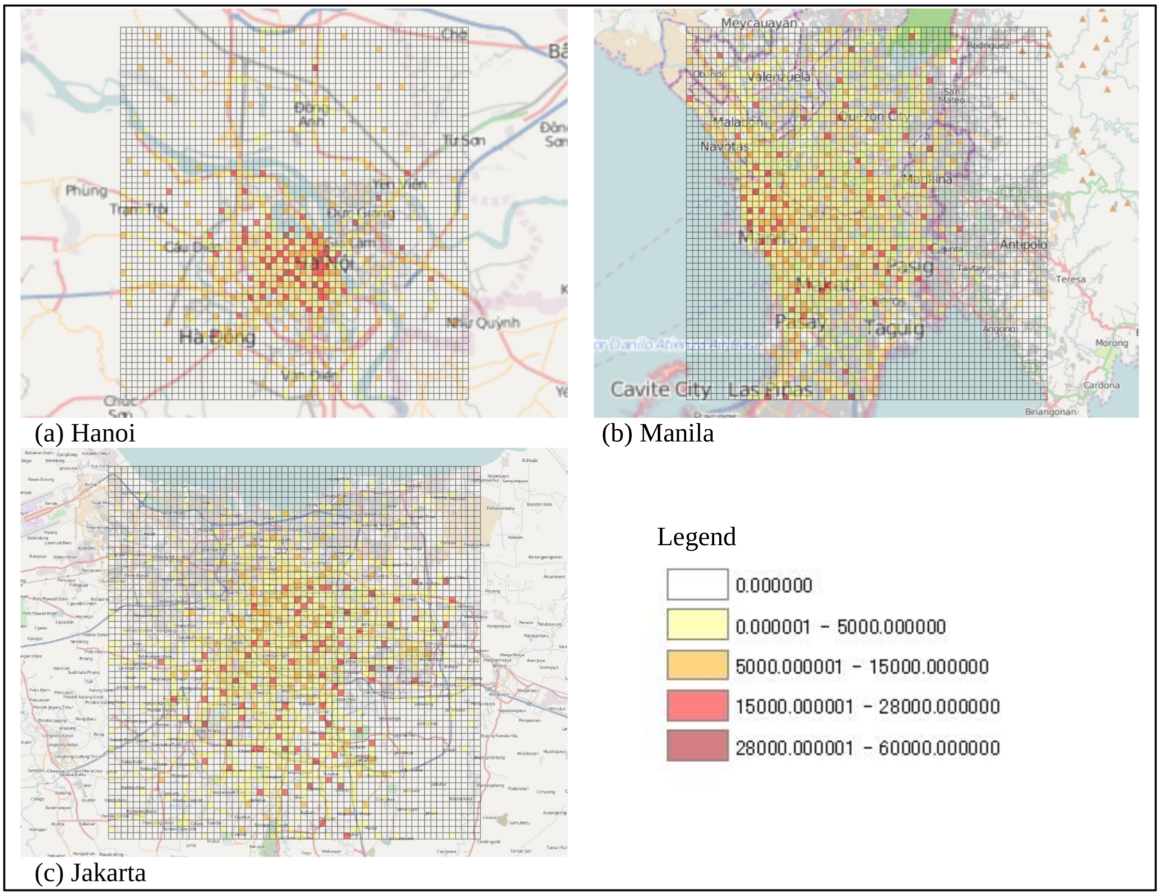


Figure 1. Population distribution at 12:00PM in three cities (a) Hanoi, (b) Manila, (c) Jakarta

3. RESULTS

The population distribution at 12:00 noon for Hanoi, Manila, and Jakarta are as seen in Figure 1. A rough comparison of the grids of three cities tells us that those in Hanoi seem to be located in the same grids, suggesting that the central area is quite congested. In comparison, the population is distributed more evenly in Manila and Jakarta, and the densely populated areas are more spread out. One problem we found was that in the more sparsely populated areas, such as around Hanoi, population tended to be concentrated in a very few number of grids.

Data discarded because they are out of range are shown as percentages of entire valid data in Figure 2. We can assume that people in Jakarta are less likely to move outside of the grid boundaries, especially compared to their counterparts in Hanoi. It is also interesting to note that those in Jakarta tend to move out of this area during the daytime, whereas the opposite could be said for those in Hanoi.

The following sections analyze results from various perspectives: focusing on the temporal changes in population distribution throughout a 12-hour period, comparing the histograms of different areas within the same city, and comparing the histograms of similar areas between the three cities.

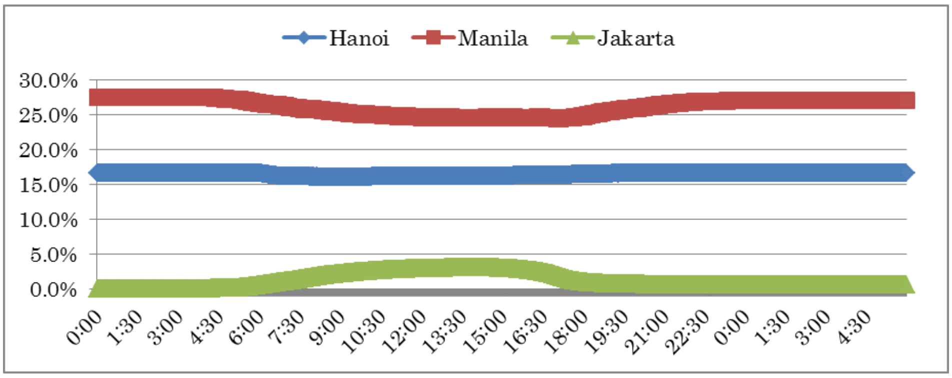


Figure 2. Percentage of total valid data discarded because out of range

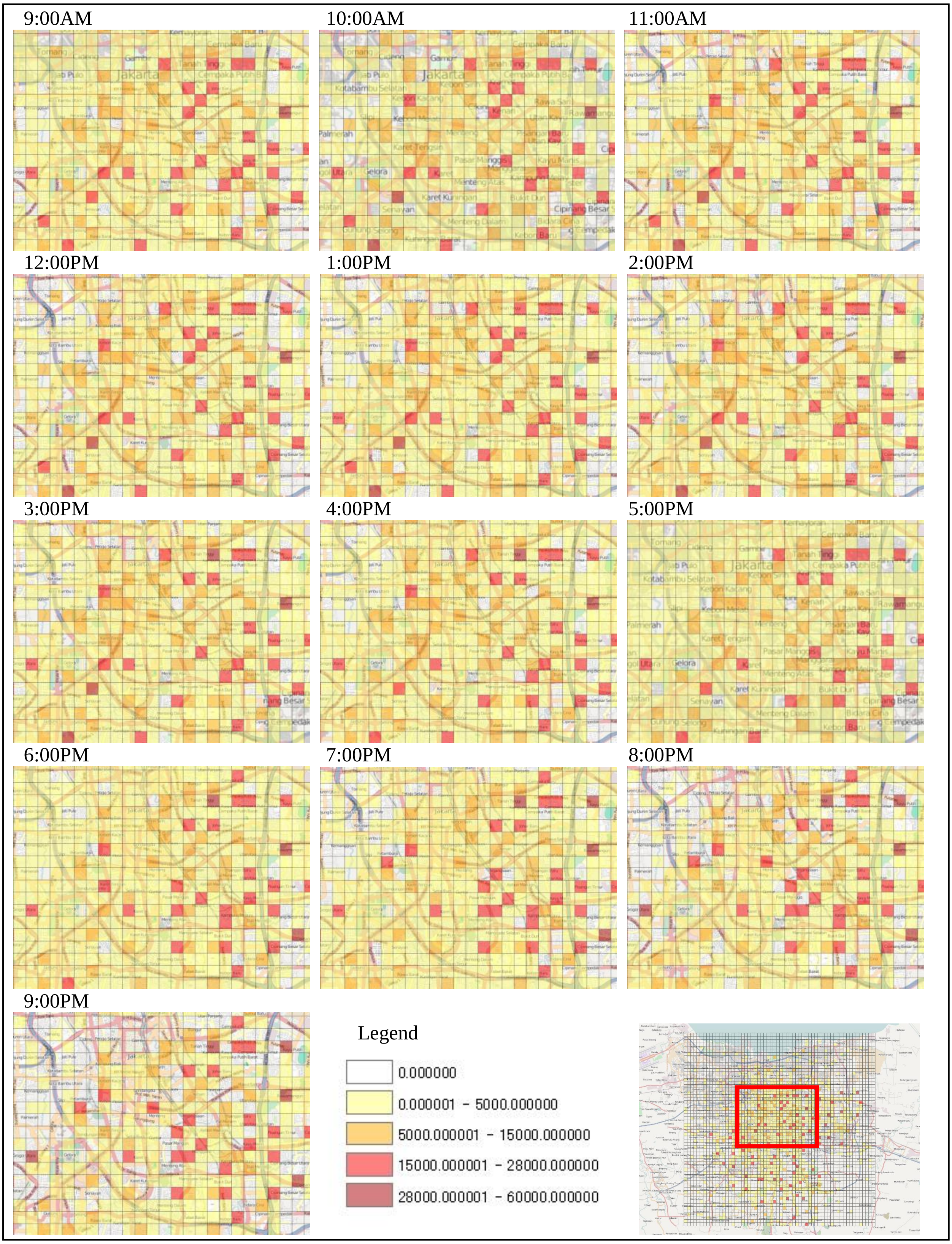


Figure 3. Population distribution in central area of Jakarta throughout the course of 12 hours

3.1 Temporal changes in population distribution

The changes in population distribution throughout the course of a 12-hour period, in the case of central Jakarta, are shown in Figure 3. We can find that certain areas, the north-western area of this block in particular, see a decrease in population density towards the evening, although different units do so at different times. This grid map can be especially helpful when focusing on the temporal changes of a single unit and the ones surrounding it.

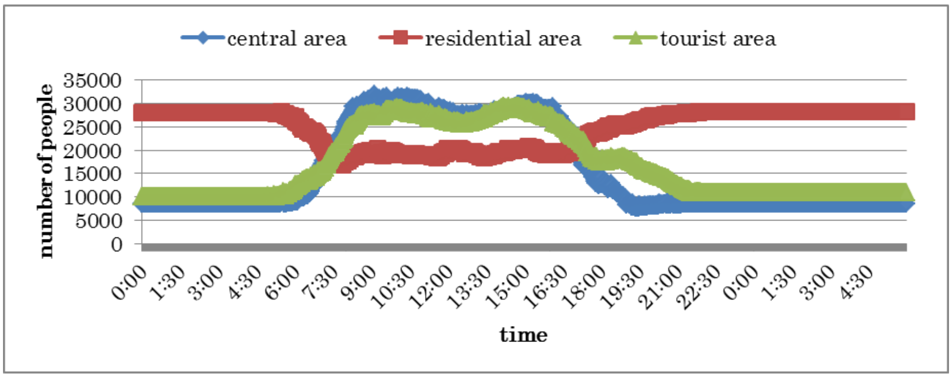


Figure 4. Histograms of different areas within Hanoi

(a) central area (b) residential area (c) tourist area

from GoogleMaps

Figure 5. Aerial photos of units used in Figure 4

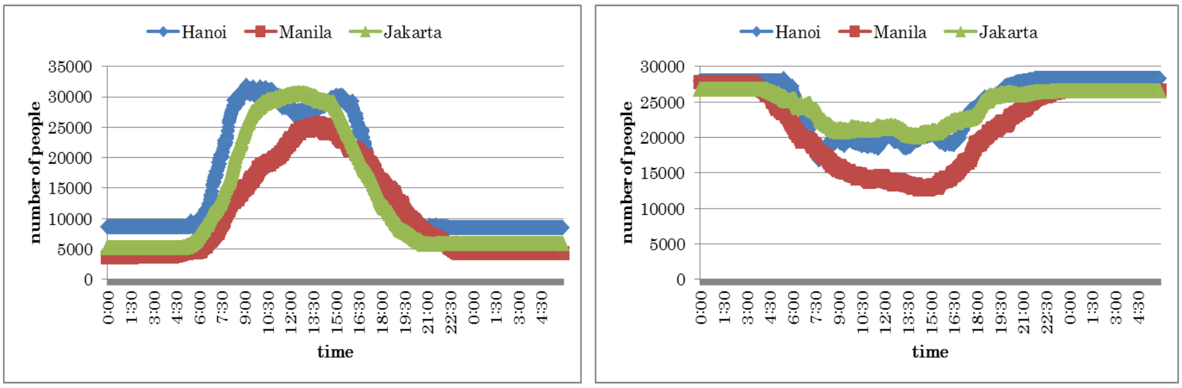
3.2 Histogram comparison of different areas within a city

Next, we compared histograms for different areas within a certain city. We made comparisons for Hanoi in units representative of a central area, a residential area, and a tourist area. The histograms and aerial photos for each area are shown in Figures 4 and 5.

From the histograms, we can confirm that during the daytime population density in central areas as well as tourist areas increases, while in residential areas it decreases. We can also assume that people in these central areas and tourist areas tend to move to a different area at around noon, presumably for lunch. Finally, when comparing the histograms for central areas and for tourist areas, we can see that tourist areas tend to stay crowded more later on into the evening, until around 9PM, whereas the population in this central area falls to its nighttime population by around 7:30PM.

3.3 Histogram comparison of similar areas in different cities

We also compared temporal changes in population distribution for similar areas within different cities. Figure 6 shows histograms making these comparisons for central areas as well as for residential areas. Regardless of city, central areas were more densely populated during the day, and the opposite could be said for the residential areas. The decrease in population density at around noon for central Hanoi was not seen in the particular examples for Manila or Jakarta. It was also interesting to note that for this particular unit in central Manila, people seem to arrive at a later time compared to those in central Hanoi and in central Jakarta. Regarding the residential areas, we can tell that those in Manila are more likely to move away from this particular unit compared to the other two cities and that in Hanoi in particular, population density in this unit fluctuates slightly throughout the afternoon.



(a) central areas (b) residential areas

Figure 6. Histograms comparing temporal changes in Hanoi, Manila, and Jakarta for (a) central areas and (b) residential areas

4. CONCLUSION

From these results, we can confirm that creating a map showing temporal changes in population distribution is indeed beneficial. We were able to examine how the concentration of people within a city varies depending on time, as well as compare changes in population density of various units throughout the course of a day. Furthermore, we were able to compare results from different cities with few problems.

One difficulty we encountered was in creating the grids. In this study we used a simple process for organizing the location data of people, but some improvements should be made for a comprehensive analysis. First of all, grid maps should cover the entire region of a city and thus, each city should have a different number of units. This would reduce the number of data that are discarded, and also reduce the task of adjusting the grid boundaries so that the majority of data would be included. Secondly, as is obvious from the grid for Hanoi, in certain areas population is concentrated in a single unit, and surrounding grids appear as though no people are located there. The reason for this is that the interpolation process used area zones that were larger in rural areas compared to urban areas, combined with the fact that these rural areas have few transportation networks. As a result, the location data of everyone in a certain zone is allocated to a single point located at the center of the zone. In order to solve this problem, unit sizes for urban areas and rural areas should be adjusted accordingly, or the interpolation method should be improved. Using a smoothing method after completion of the grid maps may be useful as well.

In the future, visualization of population distribution in cities of developing countries may be possible to produce by obtaining data from cell phones. This would enable real-time visualization of population distribution.

5. ACKNOWLEDGEMENTS

Data from person-trip surveys were provided by the Japan International Cooperation Agency (JICA).

6. REFERENCES

1. Pozzi, F., Small, C. and Yetman, G., 2003. Modeling the distribution of human population with nighttime satellite imagery and gridded population of the world. Earth Observation Magazine 12, 1.
2. Sekimoto, Y., Shibasaki, R., Kanasugi, H., Usui, T., Shimazaki, Y.: PFLOW: Reconstruction of people flow by recycling large-scale fragmentary social survey data, IEEE Pervasive Computing, (submitted)
3. http://lab.its-mo.com/densitymap/
4. LandScan: http://www.ornl.gov/sci/landscan/
5. Open Street Map: http://www.openstreetmap.org/
6. People Flow Project: http://pflow.csis.u-tokyo.ac.jp