# SPATIO-TEMPORAL ANALYSIS OF HUMAN MOBILITY IN CAIRO USING PERSON TRIP SURVEY DATA

Mohamed R. Batran [a], Hiroshi Kanasugi [c], Yoshihide Sekimoto [a], Ryosuke Shibasaki [b]

[a] Institute of Industrial Science, The University of Tokyo
(batran, sekimoto)@iis.u-tokyo.ac.jp
[b] Center for Spatial Information Science, The University of Tokyo
shiba@csis.u-tokyo.ac.jp
[c] Earth Observation Data Integration and Fusion Research Initiative, The University of Tokyo
yok@iis.u-tokyo.ac.jp

**KEY WORDS:** people flow, urban mobility. GIS analysis.

**ABSTRACT:** The rapid growth rate of urban populations may outpace the development of needed urban infrastructure, such as related to transportation, therefore, resulting to inadequacy of public transportation services and traffic congestion. Such as the case of Cairo, the capital of Egypt and is considered as an economic and political hub for North Africa. Accordingly, in order to come up with a solution to this traffic congestion problem in Cairo, initially there is need to acquire a better understanding of the urban dynamics and people flow in the capital. As such, this study aims to investigate people flow in Cairo on a macroscopic scale by utilizing person-trip (PT) survey data provided by the Japan International Cooperation Agency (JICA) in 2001. The PT survey covers daily commuting routine of a large sample size of the city population. The data was able to provide a good representation of people flow in the city despite its low temporal and spatial resolution. Pre-processing and validation of the PT dataset was initially carried out to remove errors and inconsistency. Land cover was then created so that spatial smoothing in the geocoding step can be applied in order to obtain a probabilistic distribution of the origin and destination of each trip. Using transportation network extracted from OpenStreetMap (OSM), the shortest route was determined for each trip following Dijkstra's algorithm, and finally the position of the user along the route was interpolated for one-minute intervals. The peak hours were found as 6:00-8:00 and 14:00-15:00, which correspond to the official working hours in Cairo (8:00–14:00), and the most used mode of transportation is shared taxi, followed by public bus. Finally, the output of this study may be further analyzed for planning and designing more efficient transportation infrastructure with the goal of improving urban mobility in Cairo.

## 1. Introduction

Although economic growth and urbanization of cities in the developing world supposedly benefit the city and its people, serious problems associated with such arise and need to be addressed so as not to hamper further development. The rapid growth rate of urban populations may outpace the development of needed urban infrastructure, such as related to transportation, therefore, resulting to inadequacy of public transportation services and traffic congestion. Such as the case of Cairo wherein traffic congestion costs around US$8 billion annually, which amounts to over three percent of the country's total gross domestic product, according to a recent study by the World Bank (World Bank, 2010).

Cairo is the capital and the largest city in Egypt. Greater Cairo metropolitan area is the largest urban area in Africa and the Middle East (World Bank, 2010). The population of Cairo in 2017 is at 9.5 million and is expected to increase in the future. Despite the efforts of the Egyptian government to reduce congestion by introducing new mega infrastructure projects, such as the Cairo Metro lines and Cairo ring road around the city, traffic congestion is still a serious issue imposing costs such as related to travel time, environment, and business operation. Understanding urban dynamics and human mobility is a supporting step into tailoring smart solutions suitable for the study area. Knowledge about commuting patterns, and daily spatial and temporal demand enables designing an ad-hoc type of public transportation, or applying a smart toll system to limit the usage of private vehicles in favor of public transit. In addition, studying the spatial attributes of trips such as trip length and usual transportation modes in addition to temporal characteristics such as movement peaks can give ideas about introducing an alternative transportation mode that is suitable for this case. For example, promoting cycling in areas where the trip length is short and the trip duration is high due to congestion.

Traditionally, travel survey is utilized to understand the population's travel behavior. A travel survey usually collects socio-economic and demographic attributes of households, as well as the typical daily travel diary of the interviewee, which includes information such as start and arrival times of travel, origin and destination, transportation mode, and the trip purpose. Previous studies show how PT survey data can be utilized to construct the dynamic flow of people

in the city (Sekimoto, Watanabe, Nakamura, & Horanont, 2013), (Sekimoto, Shibasaki, Kanasugi, Usui, & Shimazaki, 2011).
In this research the dynamic flow of people in Cairo was reconstructed using JICA PT survey in 2001, an ad-hoc land cover was created by classifying Landsat imagery to geo-smooth the spatial position of the origin and the destination of each trip using probabilistic reallocation, a dynamic visualization of the flow with modal share attributes was presented as a video, and a 1-hour temporal resolution 3-D density map was created.

## 2. Methodology

The study area covered all of Greater Cairo including the new communities of 10th of Ramadan City and 6th of October City with an estimated population of 14 million in the year of the analysis (2000). The study area covered the whole Cairo and Giza as well as parts of Sharqia and Qalyubia governorates (JICA, 2002).
This study uses three main datasets: i) JICA person trip survey with its associated zone data, ii) OSM road network, and iii) Landsat 8 imagery covering the study area. The City of Cairo was divided into 469 zones for the survey. A zone is the smallest spatial unit to define the origin and destination of each trip. The survey was based on home interview method with the household as the major unit and collected data such as origin and destination, start and end times, trip purpose and transportation mode, in addition to other demographic and socio-economic attributes of the household. A trip generally consists of one or more sub trips, each characterized by transportation mode, journey time, and mode transfer zone. Figure [1] shows the hierarchy of the person trip survey data. The person trip survey data collected from 52,583 households by JICA in 2001 was used, the dataset contains trip information of 137,898 people performing 268,360 trips daily. The sampling methodology of the households ensure that the dataset will give the same representation as the actual population.
OSM is a digital map database created through crowdsourcing by geographic information volunteers and is supported by the OSM Foundation. The OSM geographic information dataset for Egypt is available for download from GEOFABRIK. The extracted data on February 13, 2017 contained 1,092,160 road links and 820,239 nodes. The road links connectivity was assessed, while the disconnected links were deleted. This resulted in 1,043,221 links and 773,584 nodes with total length of 309,317 km. Lastly, Landsat 8 tiles that covers the study area were downloaded for later use in creating the custom land cover data.

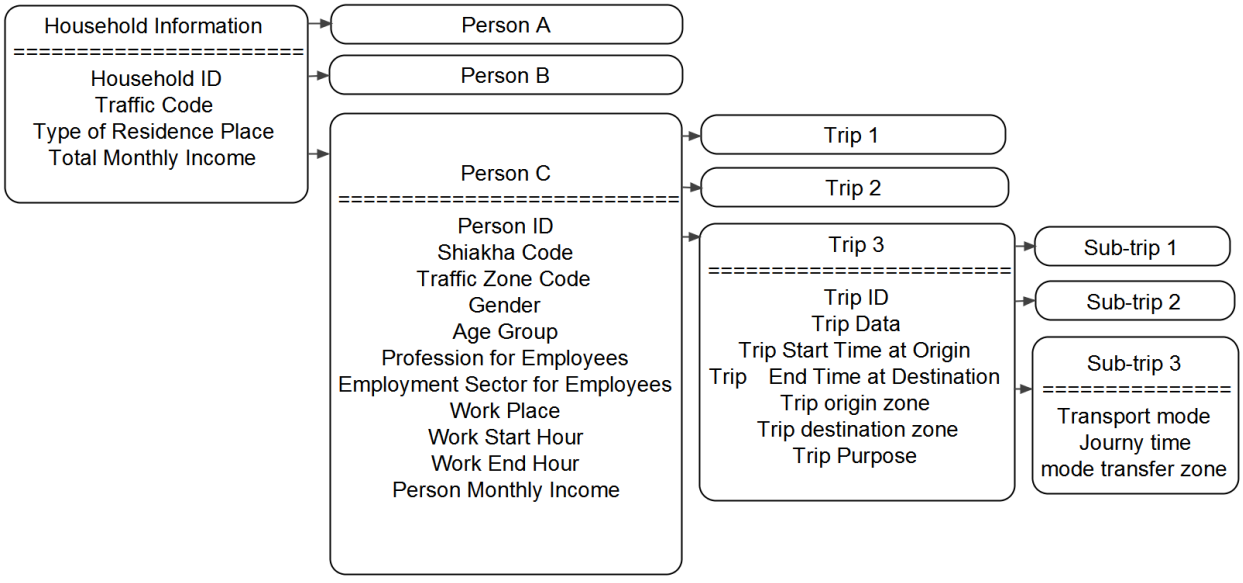


*Figure 1. PT survey data structure*

Due to the lack of appropriate land cover data, we developed our own land cover by classifying 30 square meters spatial resolution Landsat imagery using maximum likelihood supervised classification. To fully cover the area of interest, four scenes of Landsat imagery were downloaded. Since classification is sensitive to pixel intensity variation, histogram matching was applied and the four scenes were integrated into one mosaic. By checking the resulted mosaic, the land cover features in Cairo can be classified into water, urban, cropland, and desert. Training samples for each feature were collected carefully avoiding samples that are too small or too large. Very small samples might not provide sufficient information to characterize the feature, while a very large sample increase the potential of including multiple feature pixels in the same sample. Generally each training sample was maintained between 10n and 100n pixels, where n is the number of bands, i.e. 30 and 300 pixels. By evaluating the signature in the feature space, it was found out that they are well separated. The resulted classification was checked for errors, and

misclassified pixels were removed by applying the majority filter. Figure [2] shows the process of the classification. The output land cover data set has a resolution of 30 square meters.

PostgreSQL database was used to handle over one million road links. OSM road links were converted into routable network using pgRouting extension, which extends the PostGIS / PostgreSQL geospatial database to provide geospatial routing functionality.

Figure [3] represents the entire processing framework. The preprocessing stage of the PT survey data included parsing and applying some validation criteria. The validation criteria included removing records that does not match the expected range of input. For example, gender should only be 1 or 2 corresponding to male and female respectively. We also removed any duplicated users. For each trip, the origin and destination zone as well as the start and the arrival time must be known and the arrival time must be after the departure time or else the whole person id records is deleted. The validation criteria resulted in discarding 20,425 trips.

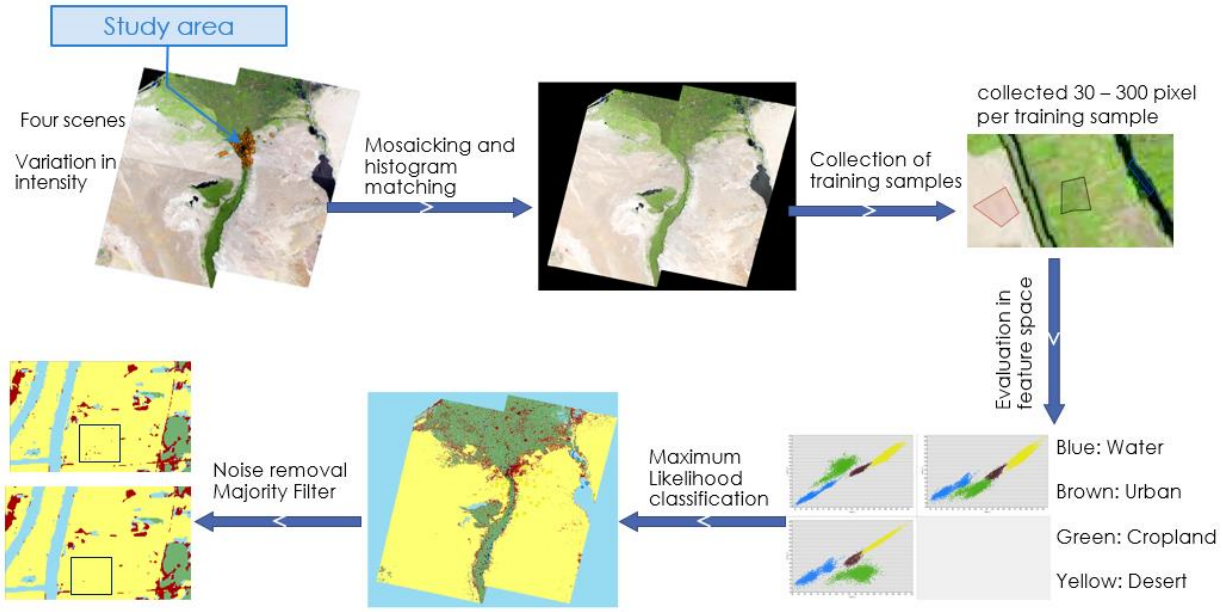


*Figure 2. Classification procedure to create land cover from Landsat imagery*

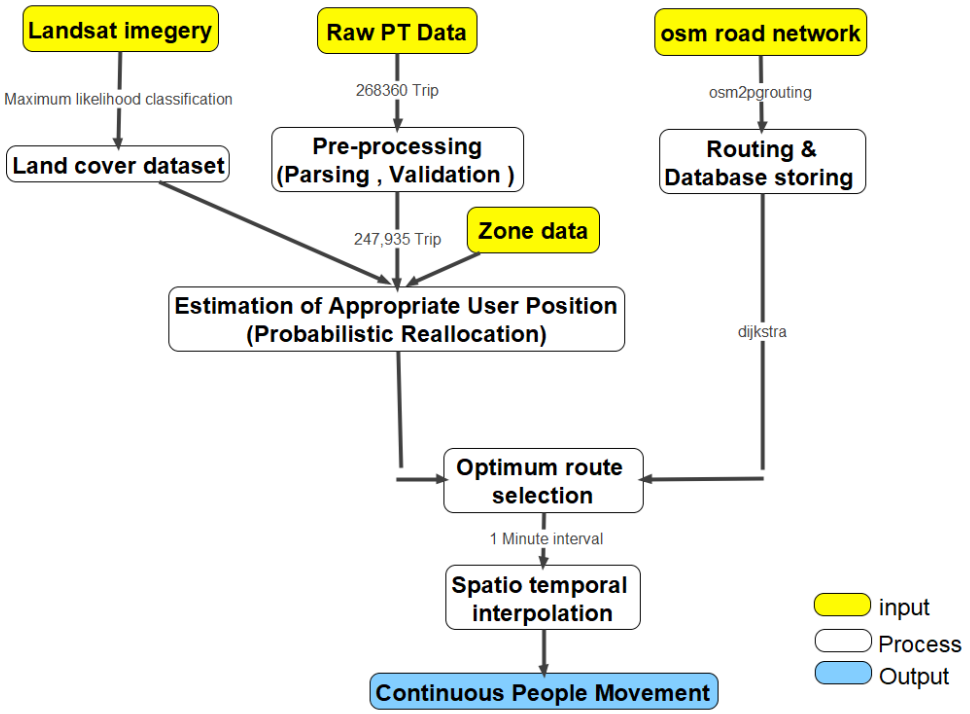


*Figure 3. Analysis framework*

The origin and the destination of each trip is given in the zone space resulting in a very low spatial distribution. In order to avoid geocoding the zone only to each center coordinates, probabilistic geo-smoothing was applied using the resulting land cover data. Each pixel is weighted according to its land cover giving a high value to urban areas, lower values for croplands and deserts, and zero for water. The sum of the total weights equal unity. The positions of each trip were geocoded on the basis of the assigned weight. This resulted in spatially smoothed population distribution. Having the spatially smoothed origin and destination coordinates of each trip, simple Dijkstra's shortest path algorithm was applied to find the optimum route. The start and arrival times were then used to interpolate the spatio-temporal position of each user along the previously selected route in 1-minute intervals. The resulting people flow dataset is the location of every sampled ID in Cairo with 1-minute temporal resolution over 24 hours.

## 3. Results and Discussion

Initial analysis of the PT survey data shows that most of the participants are aged 10-19 years old, followed by participants aged 20-29, as shown in Figure 4-b. While in terms of gender, almost two-thirds of the survey participants are male, as shown in Figure 4-c. Commuters in Cairo rely heavily on shared taxis, as shown in Figure 4-a. The same figure shows the dominance of shared taxis, which are mostly operated privately by individuals, over other public transportation modes.

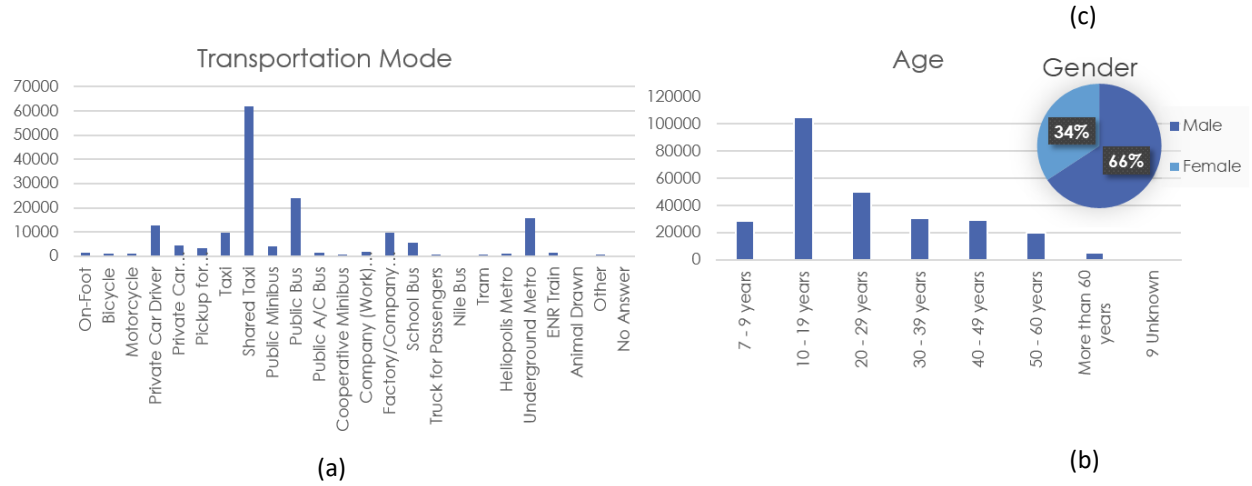


*Figure 4. (a) Modal share in Cairo, (b) age classification of PT survey data, and (c) male to female ratio of PT survey data*

By aggregating the daily travel distance per user ID, it was found out that 33 percent of people tend to travel less than 5 km daily, while 4 people out of 100 travel more than 50 km daily. It was also found out that the average travelled daily distance is 15 km for males, which is significantly higher than that of females at 9 km. In addition, two distinct peaks were observed in the hourly distribution of trips, i.e., 7:00 and 14:00. These peaks correspond to the government's official working hours (from 8:00 to 14:00), as shown in Figure 5

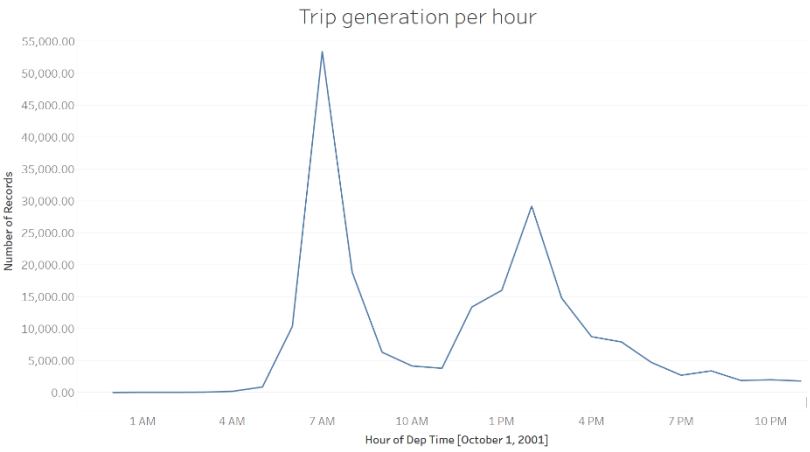


*Figure 5. Hourly distribution of the total number of trips generated*

The dynamic flow of people with their modal share during the day was then visualized using Mobmap, a visualization and analysis web application for movement data, as shown in Figure 7. In addition, Figure 6 shows the spatio-temporal distribution within the day at 1-km$^2$ mesh resolution. It can be observed that the population density of the new communities outside of Cairo, as marked with a red circle, increase during the day where most of the people work. It then starts to decrease from 18:00 when people finish work and commute to their home.

Despite the ability of the PT survey data to represent the macroscopic view of urban mobility in cities, the existence of other data sources with higher temporal or spatial resolution can lead to more applications with a better predictive accuracy. GPS logs provide better spatial resolution, and maps the exact user route with a selective temporal resolution up to seconds and is more suitable for congestion mapping and travel behavior analysis. Call detail record (CDR) is a side product of mobile network generated for billing purposes where the user location is stored in a zone spatial resolution such as the PT survey data in case of any phone activity such as making a phone call or sending a text message. CDR data is cheap to collect and can cover a larger study area of up to national level and longer periods as compared to GPS or PT survey data. Should CDR and GPS logs in Cairo become available, their accuracy for different datasets and different applications can be compared.

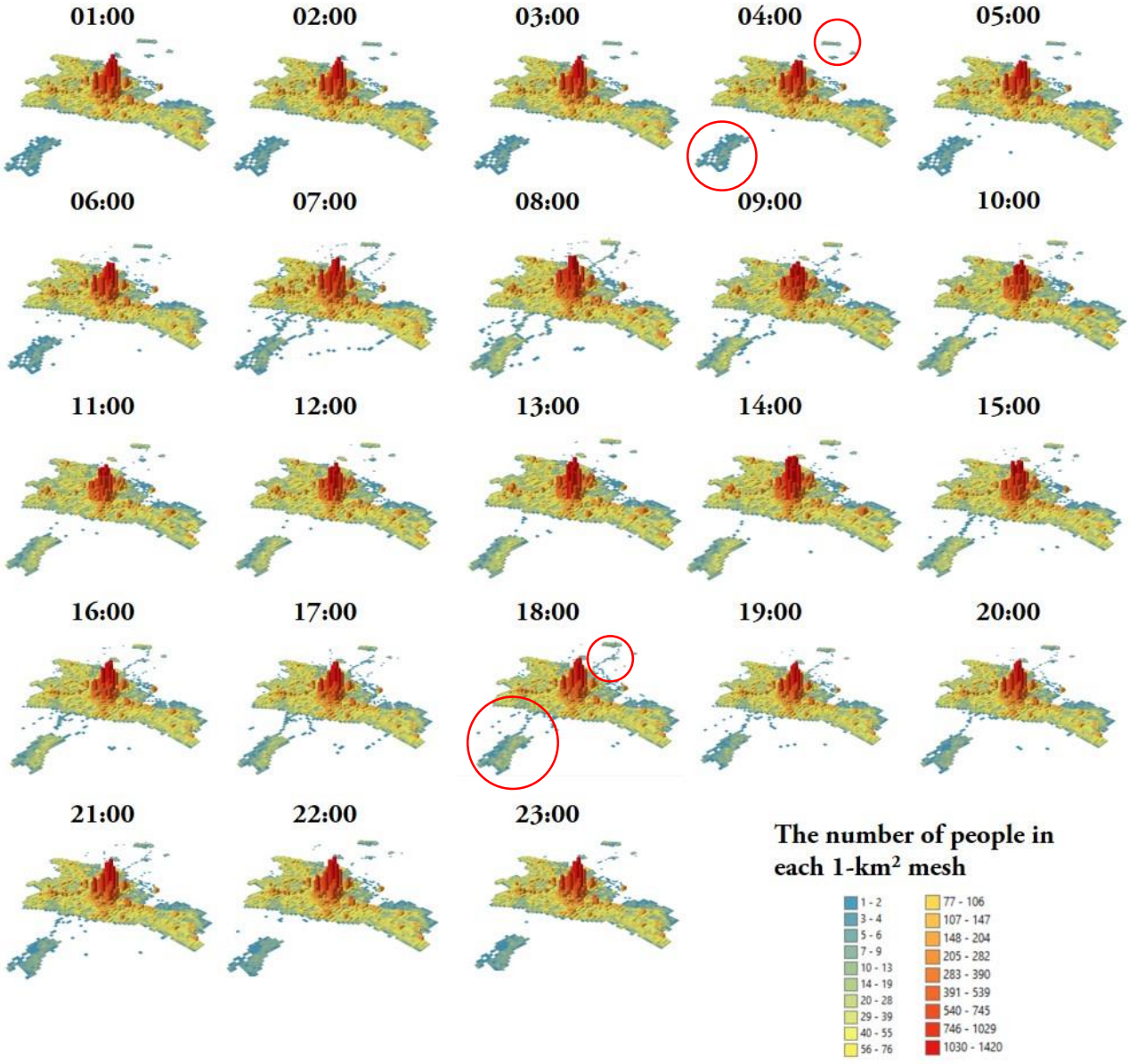


*Figure 6. 3D histogram of the number of people per 1 km$^2$ at 1-hour temporal intervals*


*Figure 7. Screenshot of the Mobmap visualization*

## 4. Conclusion and Future Work

In this research, a dynamic representation of people flow in Cairo was reconstructed using JICA PT survey data with open source infrastructure data such as OSM road links and land cover derived from Landsat imagery. A unique color for each transportation mode was assigned, and people mobility and modal share in all Cairo were visualized using the Mobmap web application. Finally, a spatio-temporal 3D density map was created identifying the density of each square km in 24 hours of the day. The results of this research provide a better visualization of the static textual person trip data for urban and transportation planning purposes. They can also be used by private businesses such as for geo-marketing as an input for selecting strategically locating advertisements and billboards.

In the future, the same methodology can be applied using the JICA 2012 PT data once it is obtain from the concerned authority. The changes in the flow of people can be compared between years 2001 and 2012, after a decade of implementing multiple mega construction projects by the Egyptian government to facilitate flow into the capital. GPS logs and CDR data, which may provide better insights with a wider range of applications, can be analyzed should they become available in the future.

## 5. References

World Bank. (2010). Cairo Traffic Congestion Study.

Sekimoto, Y., Shibasaki, R., Kanasugi, H., Usui, T., & Shimazaki, Y. (2011). Pflow: Reconstructing people flow recycling large-scale social survey data. IEEE Pervasive Computing, 10(4), 27–35.

Yoshihide Sekimoto, Atsuto Watanabe, Toshikazu Nakamura, Hiroshi Kanasugi, Tomotaka Usui, Combination of spatio-temporal correction methods using traffic survey data for reconstruction of people flow, Pervasive and Mobile Computing Journal, Elsevier, Vol.9, pp. 629-642, (Impact factor: 1.25 in 2011), 2013.

JICA. (2002). Transportation Master Plan and Feasibility Study of Urban Transport Projects in Greater Cairo Region in the Arab Republic of Egypt.