Questionnaire-based person trip visualization and its integration to quantitative measurements in Myanmar

S Kimijima1 and M Nagai2

1School of Engineering and Technology, Asian Institute of Technology, Pathumthani, Thailand
2Center for Spatial Information Science, The University of Tokyo, Tokyo, Japan

E-mail: st114307@ait.asia

Abstract. With telecommunication development in Myanmar, person trip survey is supposed to shift from conversational questionnaire to GPS survey. Integration of both historical questionnaire data to GPS survey and visualizing them are very important to evaluate chronological trip changes with socio-economic and environmental events. The objectives of this paper are to: (a) visualize questionnaire-based person trip data, (b) compare the errors between questionnaire and GPS data sets with respect to sex and age and (c) assess the trip behaviour in time-series. Totally, 345 individual respondents were selected through random stratification to assess person trip using a questionnaire and GPS survey for each. Conversion of trip information such as a destination from the questionnaires was conducted by using GIS. The results show that errors between the two data sets in the number of trips, total trip distance and total trip duration are 25.5%, 33.2% and 37.2%, respectively. The smaller errors are found among working-age females mainly employed with the project-related activities generated by foreign investment. The trip distant was yearly increased. The study concluded that visualization of questionnaire-based person trip data and integrating them to current quantitative measurements are very useful to explore historical trip changes and understand impacts from socio-economic events.

1. Introduction

The economic transition from an isolated country to opening up to the global economy is creating opportunities to develop a high potential for economic growth for Myanmar. This is mainly due to the country’s abundant natural resources, surplus labour and locational advantage, due to the geographical proximity of economic giants such as the People’s Republic of China and India. [1] has predicted Myanmar’s Gross Domestic Product (GDP) to grow by 8.3% and 8.2% in 2015 and 2016, respectively, through both government and foreign investments in the infrastructure, manufacturing and services sectors. For instance, in the telecommunication sector, mobile phone penetration is expected to increase from only about 20% now to 50% of the population in 2017. Similarly, mobile phone connectivity is expected to increase from 12% in 2014 to 70% by 2017 [2]. As the level of use and number of mobile phones users are increasing, analysing individual trip behaviour using mobile phone data can be a powerful alternative tool to capture a large volume of data and visualize the dynamics of individual trip behaviour.

In developing countries, socio-economic factors determine local life styles, including trip behaviour. Thus, evaluation of local impacts associated with socio-economic factors from a person trip
perspective is very important and Southeast Asia is not an exception. However, even though person trips are known to be affected by factors such as increased investment and improved infrastructure, changes in trip patterns cannot be traced, mainly because of the absence of established methods, the lack of secondary data and its unreliability. In the field of social science, traditional methodologies to analyse a person trip largely depend on questionnaire surveys and travel diaries. However, integrating questionnaire-based data to quantitative measurements and visualizing such trips data in time-series can be a very helpful tool to assess the diverse individual impacts resulting from rural socio-economic changes. Therefore, the objectives of this study are to: (a) visualize questionnaire-based person trip data and validate it by applying GPS loggers and (b) compare the errors between the questionnaire survey and GPS data sets with respect to sex and age group and (c) assess the change of trip behaviour. This research is expected to contribute to the visualization of questionnaire-based survey data and integrate it with quantitative measurements in order to trace dynamic individual trip behaviour by social attributes, in time-series.

2. Data
In this study, two datasets were utilized. First, a questionnaire-based person trip was prepared. A field survey was conducted with the questionnaire and a GPS device to obtain information, both for a personal profile and one-day trip as well as the house location. Second, 38 GPS loggers were distributed the respondents, two days before conducting the questionnaire survey. The trip data obtained from the questionnaires was validated by the GPS logs data. 345 individual data sets were collected from the questionnaire and GPS loggers and trip behaviour was analysed.

2.1. Study area
The study area is located in Dawei, Tanuntharyi Region, Myanmar. This region is developing its industry and infrastructure to create local employment opportunities and ultimately to support national economic growth through the Dawei Development Project which was established in 2008. The project includes development of the Dawei deep seaports, an industrial estate called Dawei special economic zone (SEZ) and road and rail links to Thailand. The Dawei SEZ law is also established to enhance the Dawei Development Project in order to be a hub for the trade and transportation in the Southeast Asian region [3].

Villages in the SEZ mainly depend on agricultural activities such as plantation and paddy cultivation; however, the project development has led to the reconstruction of villagers’ trip patterns due to the creation of employment opportunities and the development of roads. The study purposively selected rural villages dependent on agriculture and located near the Italia–Thai Development Public Company camp, which is the initial contractor with the Myanmar government for the Dawei Development Project (Figure 1.)

![Figure 1. Special economic zone in Dawei.](image)
2.2. Field survey data
A field survey was conducted in 2015 to collect information both for person trip behaviour and personal profiles. Stratified random sampling by sex and age was employed to understand trip characteristics. Non-spatial personal attributes such as age, sex, marital status, education level, household status, occupation and monthly household income were collected through a pre-tested questionnaire conducted earlier in 2014. Spatial information such as trip origin, destination, direction, distance, mode and duration in 2005, 2010 and 2015 was also collected. The study also employed formal and informal interviews with key informants such as village heads, using checklists. The data collection also involved direct field observation. The local language was used for communication with local people, assisted by a translator.

2.3. GPS log records
Wearable GPS devices such as “i-gotU USB Travel & Sports Logger – GT-600” were used to log trip data and validate the data obtained from the questionnaires. This device is lightweight (37g) and small (44 x 41.5 x 14 mm) with an automatic motion detector [4] which can be worn on the waist or clipped to clothes. The device recorded 24-hour trips with a 5-second interval using the motion detection mode. A maximum of 38 devices were distributed at one time to a total of 345 respondents aged 16 years plus. Both questionnaire and GPS log data sets are available from the 345 samples.

3. Methodology
3.1. Overall methodology
The following Figure 2 is the overall methodology employed in this study. The methodology focused on five major steps in order to achieve the main objective of integration of two data sets. First, questionnaire-based, non-spatial one-day trip data was manually converted to spatiotemporal data with the use of the respondents’ house location collected by the GPS device and visualized. Second, trip parameters such as stay points, moving segments, number of trip, trip distance and trip duration were extracted, both from the questionnaire and GPS data. Third, errors of two data sets were calculated by the selected parameters such as number of trip, trip distance and trip duration. Fourth, errors of the selected parameters were further analysed by social parameters such as sex and age. Fifth, the trip behaviour in 2005, 2010 and 2015 was analysed.

![Figure 2. Overall methodology.](image-url)
3.2. Process of questionnaire survey data
3.2.1. Conversion of questionnaire survey data to spatiotemporal data. Details of person trip obtained through interviews are ambiguous as, in rural areas, there are no significant landmarks such as prominent shopping centres and buildings as in big cities. In order to obtain a clearer picture of the trip behaviour of local people, trip information such as trip origin, destination, distance, direction and time in 2005, 2010 and 2015 was collected through the questionnaire survey. This information was manually converted to spatiotemporal information by utilizing Geographic Information Systems (GIS) and visualized in a moving format. The destination is determined using the original house location data recorded by GPS device, distance and direction. All spatiotemporal data are listed in a timeline and saved in a comma-separated values (csv) format. Simultaneously, non-spatial attributes such as age, sex, marital status, education level, household status, occupation and household income, were also integrated with the file.

3.2.2. Extraction of moving segments from questionnaire data. The total number of trip, the total trip distance and trip duration were manually extracted from the questionnaires and listed in excel format. In this study, a single trip is defined as staring from a location to a destination, such as from home to workplace. The total number of trip, distance and duration were summed for further data validation.

3.3. Process of GPS logger data
3.3.1. Stay point extraction from GPS logger data. After the trip data had been obtained with GPS loggers, spatiotemporal data such as time, latitude and longitude were extracted from the devices. The break up of the trip segment was acquired to find the stay points. In this study, stay point extraction with outlier detection and removal technique utilized by [5] were employed using the following equation (1):

\[
\text{Distance (p_{start}, p_{end})} < D_{\text{threh}} \text{ and TimeDiff} (p_{start}, p_{end}) > T_{\text{threh}}
\]  

Where the parameters D_{threh}, considerable maximum distance as a stay point, and T_{threh}, minimum time spending at a same place, are adjustable. In this study, a stay point is detected if T_{threh} > 20 minutes and D_{threh} <= 300 metres. Based on the calculation, stay points are extracted and are listed by start-time, end-time, duration, distance in metres, average speed in kilometres/hour and the total number of stay points. Additionally outlier detection and noise removal technique were applied by using standard deviation (\(\sigma\)).

3.3.2. Moving segment extraction from GPS logger data. Once the stay points are extracted from the GPS data, the moving segments can be extracted. The extracted parameters in this study is total trip distance per trip in metre, duration per trip in minutes, starting time, ending time, average speed kilometres/hour and total points. These parameters were calculated by utilizing Java language. Furthermore, the selected parameters such as the number of trip, total trip distance and total trip duration from the GPS data were summarized by utilizing the PostgreSQL.

3.4. Error calculation of two data sets
After processing the two data sets, the following equation (2) was employed to calculate the error between the two data sets by the number of trip, trip distance and trip duration.

\[
\text{Relative Change (x, y)} = \frac{\text{Absolute difference}}{\text{Max (x, y)}} \times 100 = \frac{|\triangle|}{\text{Max (x, y)}} \times 100
\]  

Where x is the data from the questionnaires and y is the data from the GPS loggers. In this process, trips made out of villages such as in Yangon and Thailand, and unfixed trips such as daily employment at various places within or outside the villages, were excluded.
3.5. Comparison of errors by social parameters
The errors were further analyzed by social parameters such as sex and age. The age groups were categorized as (A) <20, (B) 21-30, (C) 31-40, (D) 41-50, (E) 51-60 and (F) 60< for this analysis. The difference from the average error was also calculated by the categories.

3.6. Analysis of person trip behaviour
Based on the level of error, the average trip behaviour in 2005 and 2010 and 2015 obtained from the questionnaires was further analysed in time-series.

4. Result and discussion
4.1. Conversion and visualization of person trip data
For visualization of person trip data from the questionnaire, the questionnaire-based data was converted (Figure 3) to the spatiotemporal data and visualized together with GPS log data (Figure 4). This person trip can be also displayed according to attributes. This visualization of person trip data provides more useful trip characteristics by attributes than text-based information in questionnaires. Illustrative visualization is easy to understand and more information can be obtained for a better understanding of the underlying tendency behind the data [6].

4.2. Comparison of two data sets
To validate the questionnaire-based data, the average errors between two data sets were calculated. The 238 data sets from a total of 331 samples are considered in this analysis. This reveals a total of 543 trips, with a total 1,936.3 km trip distance and 7,457.0 minutes of trip duration from the questionnaires while data from the GPS loggers shows a total of 403 trips, 2,302.0 km trip distance and 7403.6 minutes of trip duration. The result shows the average errors between the two data sets as 25.5% (A), 33.2% (B) and 37.2%, respectively (Table 1).

| Average differences (%) |   |
|-------------------------|---|
| (A) No. of Person Trip  | 25.5 |
| (B) Trip Distance       | 33.2 |
| (C) Trip Time           | 37.2 |

Both human and data processing errors are found to have significantly impacted the average errors within each of the above three parameters. Human error from the questionnaires, for example, involves under-reporting of short trips and exact trip routes. Because villagers did not report stays of more than 20 minutes on the way to a destination, this introduced errors in analysing GPS log data.
Additionally, errors from the GPS data processing are largely caused because of short trips of less than 300 metres, and short stays of more than 20 minutes.

4.3. Average error of person trip by sex and age

The level of errors by sex and age and differences from the average error are summarized in the Table 2 and Table 3, respectively. The notable error in the number of trip in male age group (A) and female age group (B) are found 17.6% and -13.4% more than the average rate of 25.5%, respectively. The average number of trip in the male age group (A) shows the highest time (5.5) in the questionnaires; however that of in GPS log shows only approximately half (2.5). This probably because stay points between each trip would be shorter than 30 minutes, thus some of trips are integrated to the previous trip during processing the GPS data in this age group. While in the female age group (B) shows lowest error probably resulting from a large engagement in project-related employments, including as office staff and laboratory workers at the project camp, travelling by motorbike to the workplace. As trip that they made were only two between home and workplace, thus the potential errors during processing data are minimized. On the other hand, for the average error of trip distance, the notable errors are found in the female age group (D) as 15.6%, while female age group (B) shows the lowest error (-15.3%) of the average. The female age group (D) shows the highest error of 15.6% as respondents in this group tend to be forgetful the trip that they made. Indeed, 27.3% of respondents show more than 88.3% of error. This tendency significantly affected on the female trip duration of the age group (D). For the female age group (B), this group show the lowest error among the whole group and this can be explained due to their working pattern. On the other hand, the average error of trip duration, the male age group (A) shows the lowest error (-15.1%) probably as 50% of respondents of this group made longer trips than average of total (50.9 minute). While, female age group (C) and (D) show notable errors of 10.2% and 13.1%, respectively. For the group (C), 77.8% of trip was made less than average total time, thus short trip resulted in the higher error.

| Age  | No. of Trip | Distance | Trip duration |
|------|-------------|----------|---------------|
|      | M  | F  | M  | F  | M  | F  |
| (A) <20 | 43.0 | 25.2 | 32.0 | 41.6 | 22.6 | 36.8 |
| (B) 21-30 | 27.5 | 12.0 | 32.8 | 18.2 | 41.1 | 36.5 |
| (C) 31-40 | 27.7 | 20.4 | 36.9 | 28.4 | 29.2 | 47.9 |
| (D) 41-50 | 22.4 | 32.1 | 39.3 | 49.2 | 40.2 | 50.8 |
| (E) 51-60 | 24.1 | 29.2 | 32.8 | 31.6 | 35.7 | 32.3 |
| (F) 61< | 19.9 | 22.2 | 30.5 | 25.2 | 41.9 | 31.7 |
| Avg. | 25.5 | 33.2 | 37.2 |

Note: M and F stand for male and female, respectively. Unit is percentage.

| Age  | No. of Trip | Distance | Trip duration |
|------|-------------|----------|---------------|
|      | M  | F  | M  | F  | M  | F  |
| (A) <20 | 17.6 | -0.2 | -1.5 | 8.4 | -15.1 | 0.4 |
| (B) 21-30 | 2.1 | -13.4 | -0.8 | -15.3 | 3.4 | -1.1 |
| (C) 31-40 | 2.3 | -5.0 | 3.4 | -5.2 | -8.5 | 10.2 |
| (D) 41-50 | -3.0 | 6.7 | 5.8 | 15.6 | 2.5 | 13.1 |
| (E) 51-60 | -1.3 | 3.8 | -0.8 | -2.0 | -2.0 | -5.3 |
| (F) 61< | -5.4 | -3.2 | -3.0 | -8.4 | 4.2 | -5.9 |

Note: M and F stand for male and female, respectively. Unit is percentage.
4.4. Change in person trip

Despite of errors revealed in the previous section, the level of errors are acceptable for analysing the trip data in 2015, 2010 and 2005 obtained from the questionnaire survey. Thus, trends of parameters in 2005, 2010 and 2015 were compared and the results of yearly change are described in Table 4. The results show that row (A) does not show so much difference over a period of 10 years. Row (B) increases yearly and the changes are 1.9 (2005-2010), 2.1 (2010-2015) and 4.1 (2005-2015) times. Considering no significant change in (A), the increase in (B) is probably caused by the increased trip distance per trip. Row (C) increased in 2010; however, again decreased to similar duration to 2005 in 2015. Based on the trend of (A), (B) and (C), it can be said that the main change in trip behaviour is the increase of the trip distance and this can result from the change in trip mode such as walking, travelling by motorbike, car or other modes. It can be confirmed that as the trip distance increases, the trip mode also changes. Indeed, the motorbike mode increased 5.9 times from 2005 (9.2 %) to 2015 (54.0 %).

| Table 4. Average number of person trip, trip distance and trip duration by year. |
|-----------------|-----|-----|-----|
|                | 2005 | 2010 | 2015 |
| (A) No. of Person Trip | 2.3  | 2.1  | 3.0  |
| (B) Trip Distance (km)  | 2.3  | 4.1  | 9.2  |
| (C) Trip Duration (minutes) | 40.6 | 58.4 | 46.0 |

5. Conclusion

Person trip survey is shifting from traditional questionnaire survey based to qualitative measurements with the advance of telecommunication technology. Integration of chronological questionnaire-based trip data to GPS survey and visualizing it in time-series can be very important to evaluate chronological trip changes with consideration to socio-economic and environmental events. In this study, questionnaire-based trip data were converted to spatiotemporal data and visualized. The average error between the questionnaire and GPS data sets in the average number of person trips, trip distance and trip duration were 25.5%, 33.2% and 37.2%, respectively. Furthermore, examination of error in the trip behaviour with respect to sex and age, found smaller errors among working age females mainly employed with the project-related activities. Furthermore, questionnaire-based trip behaviour made in 2005 and 2010 were further assessed and yearly increase of trip distance was found. Thus, visualizing questionnaire-based person trip data helps in tracing old trip behaviour, and the visualized trip data can be integrated to measurements by mobile phones and GPS devices. Furthermore, such assessment with respect to social parameters in time-series will improve understanding of the dynamic impacts from socio-economic events.

Reference

[1] Asian Development Bank, “Myanmar: Economy.” [Online]. Available: http://www.adb.org/countries/myanmar/economy. [Accessed: 19-Sep-2015].

[2] World Bank, “Myanmar : Empowering People for Inclusive Growth,” 2015.

[3] A. Min and T. Kudo, “Newly Emerging Industrial Development Nodes in Myanmar : Ports , Roads , Industrial Zones along Economic Corridors,” Emerg. Econ. Corridors Mekong Reg. (BRC Res. Rep. No.8), no. 8, pp. 187–230, 2012.

[4] Mobile Action, “MobileAction Technology - i-gotU Official Site.” [Online]. Available: http://www.i-gotu.com/. [Accessed: 17-Sep-2015].

[5] A. Witayangkurn, T. Horanont, N. Ono, S. Yoshihide, and R. Shibasaki, “Trip Reconstruction and Transportation Mode Extraction on Low Data Rate GPS Data from Mobile Phone,” 2013, no. June 2015.

[6] G. Andrienko and S. Augustin, “Visual Analytics Tools for Analysis of Movement Data.”