Bridging Questionnaire Survey and GPS data for Assessing Person Trip Behavior in Time-series: A Case Study in Dawei Special Economic Zone, Myanmar

Satomi Kimijima[#1], Masahiko Nagai[*2]

[#]*School of Engineering and Technology, Asian Institute of Technology*
Moo 9 km 42 Phaholyohin Rd, Klong Nueng, Klong Luang, Pathumthani 12120, Thailand
[1]st114307@ait.asia

[*]*Center for Spatial Information Science,The University of Tokyo, Tokyo, Japan*
[2]nagaim@iis.u-tokyo.ac.jp

Abstract— **Dawei is one of the newly emerging Special Economic Zones in Myanmar where rapid socio-environmental changes are caused along with constructions of ports and road links. Evaluation of local impacts associated with such socio-environmental changes from a person trip perspective is very important. Ways of obtaining trip data can be shifted from conventional questionnaire survey to GPS survey by loggers and mobile phones in future; however, bridging two data is necessary to trace trip behavior in time-series. The objectives of this paper are to: (a) convert and visualize questionnaire-based person trip data, (b) compare the questionnaire and GPS data sets associated with trip parameters, and (c) assess trip changes in 2005, 2010 and 2015. Totally, 345 individual respondents were selected through random stratification to assess one-day trip using a questionnaire and GPS survey for each. Conversion of non-spatial trip information from questionnaires was conducted by using GIS. The results show that differences of two data sets in the number of trips, distance and duration are 25.3%, 34.9% and 38.0%, respectively. Trip distance was yearly increased. The study concluded that questionnaire data can be associated with GPS data and its visualization helps understanding underlying tendencies behind the paper-based information.**

Keywords— **Person trip, Questionnaire survey, GPS-Logger, Myanmar, GIS**

I. 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. Since establishment of the Dawei Project in 2008, industries, business and infrastructure have been developed to create local employment opportunities and ultimately to support national economic growth. The project includes development of the Dawei deep seaport, an industrial estate and highway road and rail links to Thailand. Especially, the industrial zones consist of five zones such as heavy, medium, light industry and a combination of these, with a total area of 250 km2 [1]. These zones are set as a Special Economic Zones (SEZ), which can be defined as a specific geographical regions with liberal economic laws providing special incentives for investors such as tax exemption or reduction [2]. The Dawei SEZ is expected to be the pivotal hub for better connectivity and logistics among the surrounding regions. In developing countries, socio-economic factors determine local life styles, including trip behavior. 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. Traditional ways to collect and analyze a person trip data largely depend on conventional questionnaire surveys; while GPS survey by logger and mobile phone has been also available along with the development of information and telecommunication technology. Those two data can be handled separately in different field of studies such as social science and engineering; however, bridging questionnaire survey data to GPS survey data 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) convert and visualize questionnaire survey person trip data, (b) compare the questionnaire and GPS data sets associated with trip parameters such as the number of trip, trip distance and trip duration and (c) assess trip changes in 2005, 2010 and 2015 associated with social parameters such as sex and age. The research is expected to contribute to the conversion and visualization of questionnaire survey data and bridge it with quantitative measurements in order to trace dynamic individual trip behavior by social attributes, in time-series.

II. 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 to sample respondents, two days before conducting the questionnaire survey and trip data was validated by the GPS log data. 345 individual data sets were collected from the questionnaire and GPS survey and trip behavior in 2005, 2010 and 2015 was analyzed.

A. Study Area

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 after the project establishment due to the creation of employment opportunities and development of roads infrastructure. The study purposively selected rural villages dependent on agriculture and located near the Italia–Thai Development Public Company (ITD) camp, which is the initial contractor with the Myanmar government for the Dawei Project (Fig. 1).

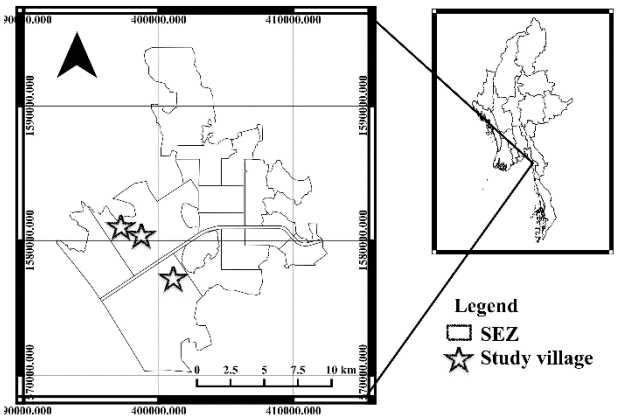

Fig. 1 SEZ IN DAWEI

B. Field Survey Data

A field survey was conducted with a questionnaire and a GPS in 2015 in order to collect information for person trip behavior and personal profiles. Stratified random sampling by sex and age was employed to understand trip characteristics and 345 individual data were collected. Nonspatial 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.

C. GPS Logger Data

Wearable GPS devices such as "i-got U USB Travel &
Sports Logger – GT-600" were used to log trips 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 [3] 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 are available from the 345 samples. The samples comprise age groups: 16-20 (16.9%), 21-30 (16.0%), 31-40 (16.6%), 31-50 (18.1%), 51-60 (19.3%), and above 61 years (13.0%).

III. METHODOLOGY

A. Overall Methodology

The following Fig.2 is the overall methodology employed in this study. The methodology focused on four major steps in order to achieve the main objective of extracting rapid socio-environmental changes. First, questionnaire-based, non-spatial one-day trip data with the respondents' house location collected by the GPS device was manually converted to spatiotemporal data. Second, trip parameters such as stay points, moving segments, number of trip, trip distance and trip duration were extracted both from the questionnaire and the GPS data. Third, the questionnaire person trip data for 2015 was validated using the GPS logger data, based on the selected parameters such as number of trip, trip distance and trip duration. Fourth, the trip behavior in 2005, 2010 and 2015 was analyzed based on socio-economic parameters of the respondents.

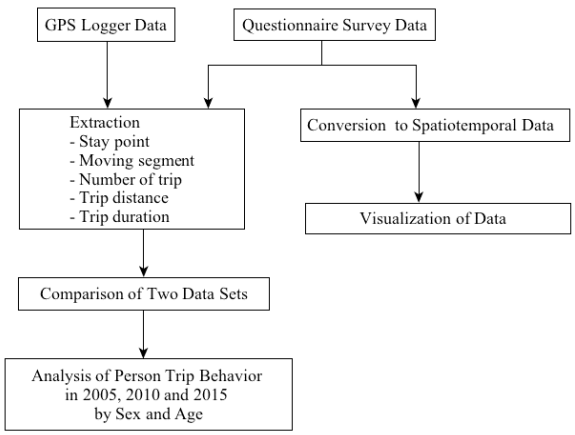

Fig. 2 OVERALL METHODOLOGY

B. Process of Questionnaire Survey Data

1) *Extraction of Trip Data from the Questionnaires:* Details of person trip obtained through interviews are ambiguous and its changes are also not clear. In order to obtain a clearer picture of local people's trip behavior, trip information such as trip origin, destination, direction, distance, and duration in 2005, 2010 and 2015 were collected. This information was converted to spatiotemporal information by utilizing Geographic Information Systems (GIS) and visualized in a movie format. The destination is determined using the original house location data recorded by the 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.

2) Extraction of Moving Segments from the 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 from a staring from a

location to a destination, such as from home to a workplace.

C. Process of GPS Logger Data

1) Stay Point Extraction from the 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 [4] were employed using the following Eq. (1):

$$Distance\ (p_{start}, p_{end}) < D_{threh}\ and\ TimeDiff\ (p_{start}, p_{end}) > T_{threh} \quad (1)$$

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 (σ).

2) Moving segment extraction from the GPS 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 are calculated for a single trip by utilizing Java language. For data reliability, duration < 1 minute and total point < 5 were excluded from the list. The selected parameters from the GPS data were summarized by utilizing the PostgreSQL.

D. Comparison of Two Data Sets

After processing the data, the following Eq. (2) was employed to calculate differences in the number of one day trip, trip distance and trip duration between the two data sets.

$$Relative\ Change\ (x, y) = |\Delta| / Max\ (x, y) * 100 \quad (2)$$

Where x is the data from the questionnaire and y is the data from the GPS loggers. Accordingly, changes in these parameters in 2005, 2010 and 2015 were calculated. 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.

E. Analysis of Person Trip Behavior

The trip distance was analyzed by social parameters such as sex and age. Furthermore, the increase rates of 2005-2010 and 2010-2015 by the parameters were compared.

IV. RESULT AND DISCUSSION

A. Conversion and Visualization of Questionnaire Data

For visualization of person trip data from the questionnaire, the converted data was visualized together with GPS log data (Fig.3). Person trip can be also displayed according to attributes (Fig. 4). This visualization of person trip provides more useful trip characteristics such as trip distance, number of trips and trip duration by parameters than text-based information obtained from questionnaires. Illustrative visualization is easy to understand and more information can be obtained for a better understanding of the underlying tendency behind the data [5].

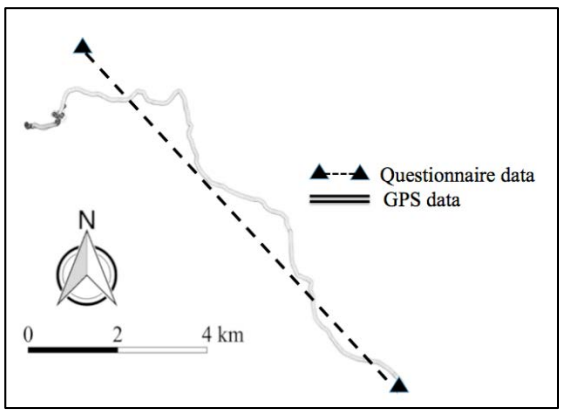

Fig. 3 COMPARISON OF QUESTIONNAIRE AND GPS DATA

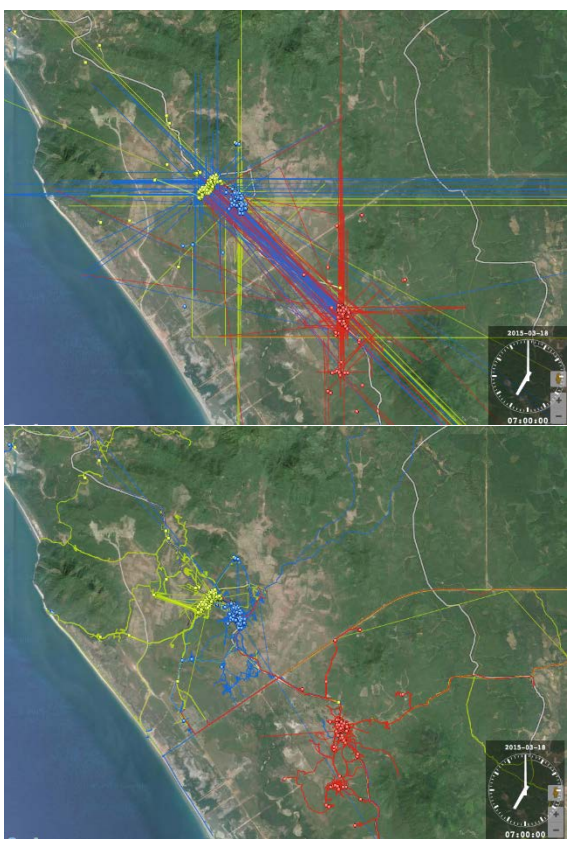

Fig. 4 COMPARISON OF VISUALIZED QUESTIONNAIRE AND GPS DATA

B. Comparison of Two Data Sets

To validate the questionnaire survey data, the average differences between two data sets were calculated. 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 questionnaire 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 differences between the two data sets as 25.3% (A), 34.9% (B) and 38.0%, respectively (Table I).

TABLE I
AVERAGE DIFFERENCE OF TWO DATA SETS

	Average differences (%)
(A) No. of Person Trip	25.3
(B) Trip Distance	34.9
(C) Trip Time	38.0

Both human and data processing errors are found to have significantly impacted the average differences within each of the above three parameters. Human error from 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 during the data processing of GPS logs data are largely caused because of short trips of less than 300 metres, and short stays of more than 20 minutes. However, the level of error allows us to utilize the trip data in 2015, 2010 and 2005 obtained from the questionnaire survey.

C. Change in Person Trip

Trends of parameters in 2005, 2010 and 2015 were compared and the results of yearly change are described in Table II. The results show that row (A) does not show 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. Row (C) increased in 2010; however, in 2015, it again decreased to a duration similar to 2005. Based on the trend of (A), (B) and (C), it can be said that the main change in trip behavior 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 II
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

D. Change in Trip Behaviour by Sex and Age

Sex and age are the important parameters for assessing trip tendencies. The trip distances in 2005, 2010 and 2015 were compared with respect to sex and age (Table III & Table V). The male trip distance during 2005-2010 and 2010-2015 increased by 3.6 and 1.7 times and that of females by 3.3 and 2.9 times, respectively for the age group (A). The notable increase in the female trip distance during 2010-2015 is due to occupational changes from being a student to working in agriculture, non-agriculture and other activities. Furthermore, some females travel to a city university, so making the search for higher education is one of the main driving factors for this age group. While in age group (B), the male rate during 2005- 2010 and 2010-2015, increases by 1.2 and 1.9 times, respectively and that of females by 1.5 and 4.9 times, respectively. The higher rate of increase for females during 2010-2015 is a result of their engagement in project-related employment, including as office staff and as laboratory workers at the project camp, travelling by motorbike to the workplace. On the other hand, for age group (C), the male rate during 2005-2010 and 2010-2015 increases by 3.0 and 1.3 times, respectively and that of females by 0.5 and 3.7 times, respectively. The lower increase for males during 2010-2015 compared to females, is probably due to non-significant changes in occupation for males. Furthermore, the higher increase for females during 2010-2015 is probably the result of increased trips to the city for business work. This tendency in age group (C) can be also found in age group (D), showing increases of 3.3 (2005-2010) and 1.2 (2010-2015) times for males and of 1.2 (2005-2010) and 2.8 (2010-2015) times for females. For age group (E), the rate during 2005-2010 and 2010-2015 increase by 1.5 and 3.0 times, respectively while that of females increase by 1.3 and 0.4 times, respectively. The notable increase in male rates during 2010-2015 is the result of increased engagement in non-agricultural activities, such as driving car taxis. Indeed due to business expansion, one male bought a car and own taxi driving service. The decreased trip distance for females during 2010-2015 is the result of non-significant changes in occupation and trip mode. For age group (F), the male rates during 2005-2010 and 2010-2015 increased by 3.0 and 1.4 times, respectively while that of females increased by 1.7 and 1.6 times, respectively. The increases in the male trip distances during 2005-2010 can be the result of increased use of motorbike mode without significant changes in occupation. While there is no significant change for females due to non-significant changes in occupation and trip mode for them.

Based on the findings of this study, it can be said that occupation and trip modes are significant factors influencing the trip behaviour of the villagers. Especially, there is significant increased engagement in 2015 in project-related occupations by females in the age group (B) and in nonagricultural activities by females in the age group (C) combined with a significant increase in use of motorbike mode of travel to work. This is consistent with the findings of [6] in rural periphery Chennai, India where employment opportunities and accessibility to work places play an important roll to determine trip

behaviour, increased working opportunities especially among working age women results in increased of travel distance.

TABLE III
AVERAGE TRIP DISTANCES BY SEX AND AGE GROUP

Age	2005		2010		2015	
	M	F	M	F	M	F
(A)16-20	1.2	1.2	4.3	4.0	7.3	11.6
(B)21-30	4.8	1.1	5.9	1.7	11.1	8.3
(C)31-40	3.0	2.2	8.9	1.1	11.9	4.1
(D)41-50	3.4	2.2	11.1	2.7	13.4	7.5
(F)51-60	4.01	3.1	6.1	4.1	18.3	1.7
(G)61<	1.4	0.9	4.2	1.5	5.7	1.6
Average	2.8	1.8	6.5	2.8	1.6	6.8

Note: M stands for male and F stands for female Unit is km

Table V
INCREASE RATE OF TRIP DISTANCE

Age	2005/2010		2010/2015	
	M	F	M	F
(A)16-20	3.6	3.3	1.7	2.9
(B)21-30	1.2	1.5	1.9	4.9
(C)31-40	3.0	0.5	1.3	3.7
(D)41-50	3.3	1.2	1.2	2.8
(F)51-60	1.5	1.3	3.0	0.4
(G)61<	3.0	1.7	1.4	1.1
Average	2.6	1.6	1.7	2.6

Note: M stands for male and F stands for female

V. CONCLUSION

The establishment of Dawai project is the significant key factor which reconstruct trip patterns of local people. Development of information and telecommunication technology and increased distribution of GPS equipped mobile phones may provide an alternative way to collect a large volume of trip information in future. However, bridging conventional survey and GPS data and visualizing them are necessary to trace trip behavior and access local impact in time-series. In this study, questionnaire survey trip data were converted to spatiotemporal information and visualized. Comparison of questionnaire and GPS data sets was conducted to find out reliability of questionnaire data. The average difference between the two data sets in the average number of person trips, trip distance and trip duration were 25.3%, 34.9% and 38.0%, respectively. Furthermore, examination of 2005-2015 trends with respect to sex and age, found large increases in motorbike-based trip distance for working age females employed either with the project or in non-agricultural activities. Thus, visualizing questionnaire survey person trip data helps to trace past trip behaviour, and the visualized trip data can be associated with GPS data. Furthermore, such assessment with respect to social parameters in time-series will improve understanding of the dynamic individual impacts.

REFERENCES

[1] A. Min and T. Kudo, "Newly Emerging Industrial Development Nodes in Myanmar : Ports, Roads, Industrial Zones along Economic Corridors," *Emerg. Econ. Corridors Mekong Reg*, vol. 8, pp. 187–230, 2012.

[2] M. ISHIDA, "Special Economic Zones and Economic Corridors," Economic Research Institute for ASEAN and East Asia (ERIA), Working Paper d019, 2009.

[3] "Mobile Action," *Mobile Action Technology - i-gotU Official Site*. [Online]. Available: http://www.i-gotu.com/.

[4] 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," in *International Conference on Computers in Urban Planning and Urban Management (CUPUM 2013)*, 2013, vol. 5

[5] G. Andrienko, N. Andrienko, and S. Wrobel, "Visual Analytics Tools for Analysis of Movement Data," *SIGKDD Explor. Newsl.*, vol. 9, no. 2, pp. 38–46, Dec. 2007.

[6] S. Srinivasan and P. Rogers, "Travel behavior of low-income residents: studying two contrasting locations in the city of Chennai, India," *Journal of Transport Geography*, vol. 13, no. 3, pp. 265–274, Sep. 2005.