Interactive Online Micro-spatial Population Analysis based on GIS Estimated Building Population

Authors Contact Information

Ko Ko Lwin
Division of Spatial Information Science,
Graduate School of Life and Environmental Sciences,
University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8572 Japan.
e-mail: kokolwin@geoenv.tsukuba.ac.jp

Yuji Murayama
Division of Spatial Information Science,
Graduate School of Life and Environmental Sciences,
University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8572 Japan.
e-mail: mura1@sakura.cc.tsukuba.ac.jp

ABSTRACT

Spatial distribution patterns of population is fully depend on landscape structures and never be a homogeneous, especially where the city has a mix of high and low-rise buildings or patched with unpopulated large spaces such as paddy fields or parks or playgrounds or governmental institutions. This will introduce some errors in population data analysis at micro-scale level. In order to eliminate these errors, we need to estimate population at building level. Spatial analysis functions using building population data is absolutely rare or absent in GIS arena because building population information is not available for public use due to privacy concerns. The goal of this paper is to introduce an online interactive micro-spatial population analysis based on building population, which was estimated by LIDAR derived Digital Volume Model (DVM) and number of floors attribute information with census tracts.

INTRODUCTION

A little attention is focused on population data analysis at micro-scale level due to lack of population data available at smallest geographical unit. Moreover, current population data used in GIS are represented as a point or polygon or grid with aggregated values. We always assumed population data as a homogeneous plane. Openshaw (1992) identified the following sources of error: errors in the positioning of objects, errors in the attributes associated with objects, and errors in modeling spatial variation (e.g. by assuming spatial homogeneity between objects). Under the GIS analysis, spatial analysis functions performed within the census tract do not acquire any significant changes in population. In order to improve the accuracy in population data analysis at micro-scale level, we need a quantitative estimate of building population and exact locations of their distributions. However, building population information is not available for use in the public domain due to privacy concerns. Nowadays, it is possible to generate population data at the building level due to the emergence of commercial remote sensing data available at finer spatial resolution with more diverse geo-information sources (IKONOS, QuickBird, LIDAR, etc.) and GIS data available at smaller map scale with better attribute information (building footprint data with number of floors, building name, block number, etc.).

On the other hand, emergence of user friendly web GIS technologies like Google Map / Earth and Microsoft Virtual Earth empower the internet users to make interactive spatial decision in timely manner. Nowadays, web GIS is part of our daily lives. For example, finding a restaurant, browsing our neighborhoods, choosing a place where to live is an important consideration for potential home buyers through the internet. Retail shops owners are ever seeking for populated areas inside the city. Urban planners are always watching for population growth and their spatial distribution patterns in order to maintain and improve the urban system.

BACKGROUND

Common cartographic forms of population mapping are the choropleth map and dasymetric map. Choropleth maps provide an easy way to visualize how a measurement varies across a geographic area. However, choropleth maps have limited utility for detailed spatial analysis of population data, especially where the population is concentrated in a relatively small number of villages, towns and cities. One way to avoid this limitation is to transform the administrative units into smaller and more relevant map units through a process known as dasymetric mapping (Bielecka 2005). There are several methods to estimate and map the population data in GIS, for example, point interpolation, areal interpolation and dasymetric mapping. Point interpolation includes a variety of methods to estimate the values. Many commercial GIS products provide point interpolation using Inverse Distance Weighting (IDW), SPLINE and Kriging methods. Areal interpolation estimates the values in target zones using the known values from a source zone, for example, census tracts to electoral zones or census tracts to watershed zones. Recent research suggests that dasymetric mapping can provide more accurate small-area population estimates than many areal interpolation techniques that do not use ancillary data (Mrozinski and Cromley 1999; Gregory 2002). U.S. Geological Survey (USGS) researchers have refined and extended automated procedures for improving spatial accuracy and visualization in mapping population distribution using “dasymetric” mapping. This technique aims to refine the spatial accuracy of aggregated data by using ancillary information to partition space into zones that better reflect the statistical variation in population (Sleeter 2008).

On the other hand, the use of Light Detection and Ranging (LIDAR) data in urban applications is increasing by means of realistic visualization of various landscapes for telecommunication and mobile network planning and identification of the relationship between urban morphology and surface temperature (Nichol 1996). Hug (1997) stated that laser scanners are the best choice for obtaining digital surface models, especially for dense urban areas. Haala and Brenner (1997) reported on similar work using airborne LIDAR data for the generation of 3D city models. Kim et al. (2000) provided a concise examination of using photogrammetric imagery and LIDAR data for obtaining a DTM in urban areas. However, incorporating LIDAR data with demographic data is very rare in GIS applications.

ONLINE INTERACTIVE MICRO-SPATIAL POPULATION ANALYSIS

Study Area

The study area is part of Tsukuba City, a planned city for academic and scientific purposes and the home of Tsukuba University and Japan Aerospace Exploration Agency (JAXA).

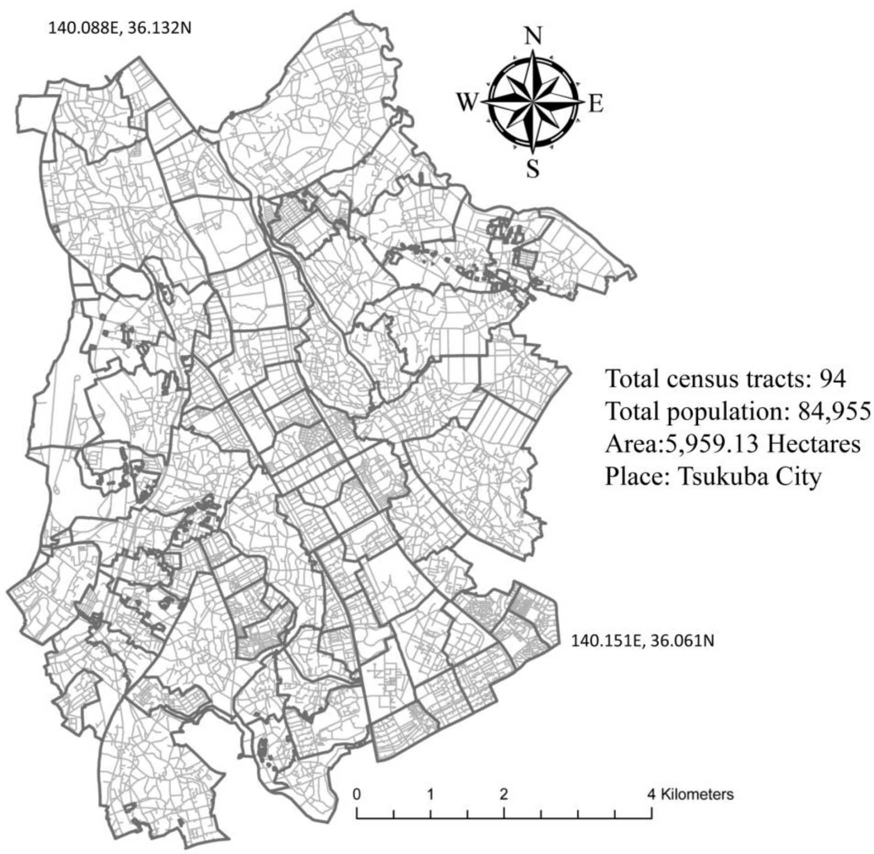


Figure 1. Study Area and Transportation Network

GIS Estimated Building Population Data

We have studied estimation of building population from LIDAR derived Digital Volume Model (DVM) and census tracts in Tsukuba City. We proposed following mathematical equations to estimate building population using GIS technique based on available attribute information.

Areametric Method

$$BP_i = \left(\frac{CP}{\sum_{i=1}^{n} BA_i}\right) BA_i$$ Using building footprint surface area (1)

Volumetric Method

$$BP_i = \left(\frac{CP}{\sum_{i=1}^{n} BA_i . BF_i}\right) BA_i . BF_i$$ Using number of floors information (2)

$$BP_i = \left(\frac{CP}{\sum_{i=1}^{n} BA_i . BH_i}\right) BA_i . BH_i$$ Using average building height (3)

$$BP_i = \left(\frac{CP}{\sum_{i=1}^{n} BV_i}\right) BV_i$$ Using total building volume (4)

Where:

BP_i	Population of building *i*
CP	Census tract population
BA_i	Footprint area of building *i*
BF_i	Number of floors of building *i*
BH_i	Average height of building *i* (from LIDAR)
BV_i	Total volume of building *i* (from LIDAR)
i	Index
n	Number of buildings that met user-defined criteria and fall inside the CP polygon

Areametric method is suitable for low-rise building areas (i.e., rural areas) where the volumetric method is suitable for high-rise building areas (i.e., downtown areas).

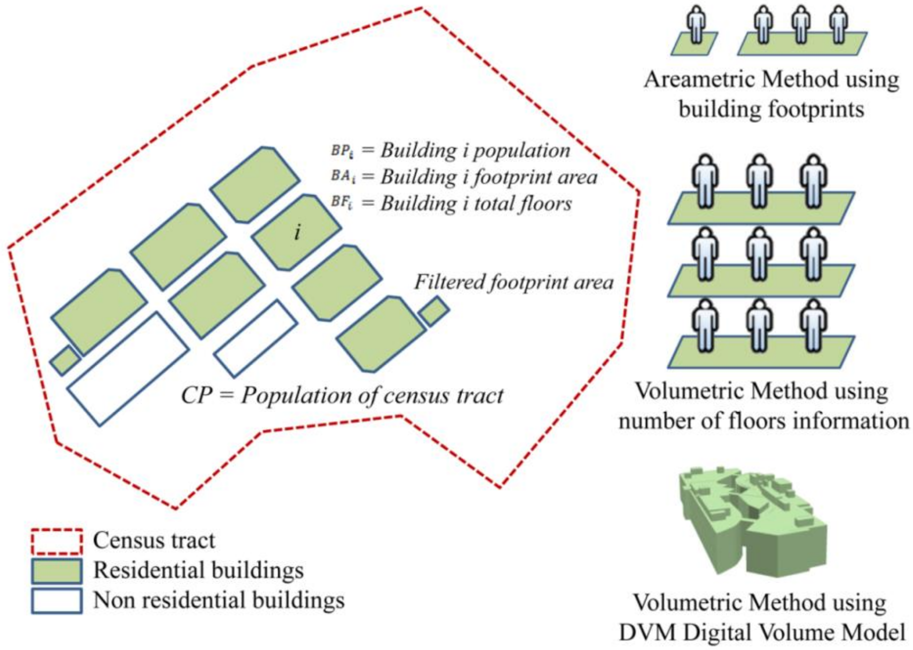


Figure 2. Graphical illustration of equations

Data Processing

In our previous study, we used LIDAR data in point format provided by PASCO Corporation and building footprints data from Zmap-TOWNII dataset which was purchased from Zenrin commercial GIS data vendor. We extracted Digital Height Model (DHM) from LIDAR data by subtracting Digital Surface Model (DSM) and Digital Terrain Model (DTM). Digital Volume Model (DVM) was computed from DHM by multiplying each pixel value with cell surface area. DHM and DVM raster data were masked with residential building footprints shape file and then extracted average building height and total volume using zonal statistical analysis function in ArcGIS. Zmap-TOWNII data also included number of floors attribute information. We used mean value of number of floor and DVM approach results. Extraction of residential footprints includes four steps: first, remove all building footprints which fall inside the zero census tracts (i.e., industrial zones, institutional areas, etc.), second, remove all building footprints which intersect with large scale business activities points which was acquired from Nippon Telegraph & Telephone Corp. (NTT) Townpage (i.e., financial institutions, governmental organizations, etc.), third, remove all building footprints heights less than 2m and surface area less than 20m^2

(i.e., garbage boxes, bicycle stand roofs, portico roofs, etc.), finally, remove such building footprints (i.e., multi-story garages, elevator section, etc.) by manually which cannot detect by automatically was using 8cm Ortho image and ground truth survey. Additional field survey and ground truth verification were required for the adjustment factor. This included, for example, measurement of occupancy and observation of mixed building-use type. Calculation of adjustment factor is based on the following equation:

AF_i = Total Number of Residential Floors / Total Number of Building Floors

For example, the adjustment factor for a four-floor building with one commercial-use floor is 0.75 (3/4) and zero for an empty building (abandoned house) or newly constructed buildings.

$$BP_i = \left(\frac{CP}{\sum_{\mathrm{i=1}}^{\mathrm{n}} BV_i.AF_i}\right) BV_i.AF_i \qquad (5)$$

AF_i Adjustment Factor of building i

Mixed building use type can be easily detected from NTT Townpage point data by intersecting home-based or small business category. We have also developed a standalone GIS tool (PopShape GIS) for estimation of building population which can be downloaded from following URL (http://giswin.geo.tsukuba.ac.jp/sis/en/software.html). Final estimated results were validated by single-unit apartments with multiple rooms (normally referred to as 1K apartments, one room with kitchen) by confirming occupancy through mailbox usage.

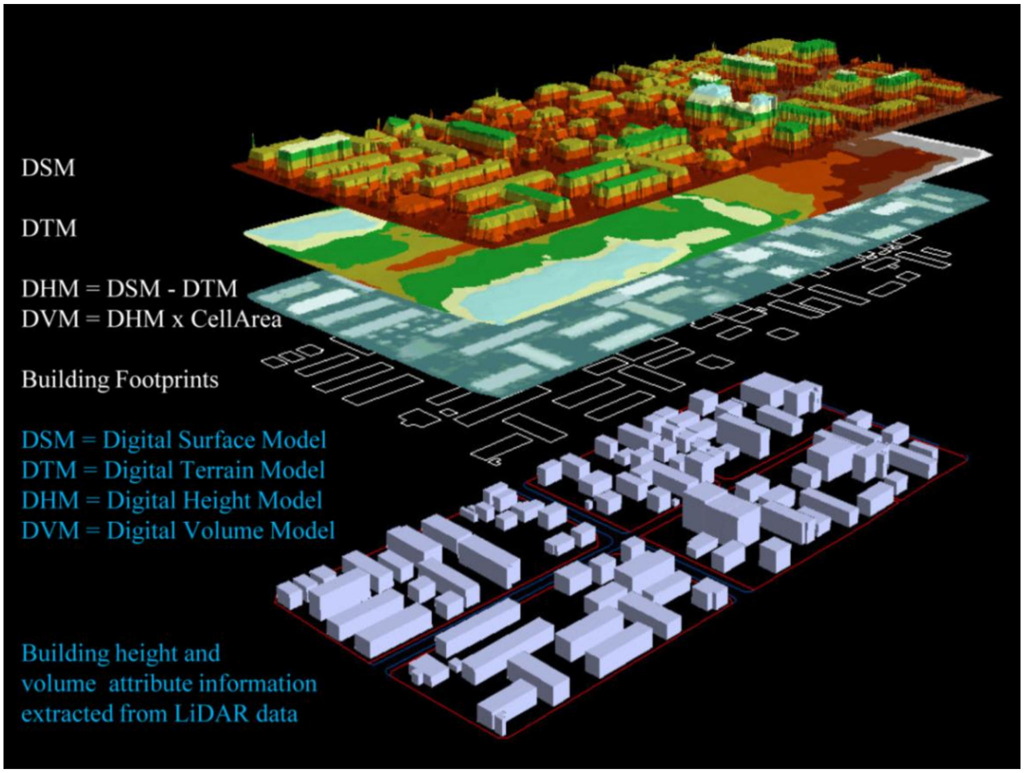


Figure 3. Extraction of average building height and total volume from LIDAR data

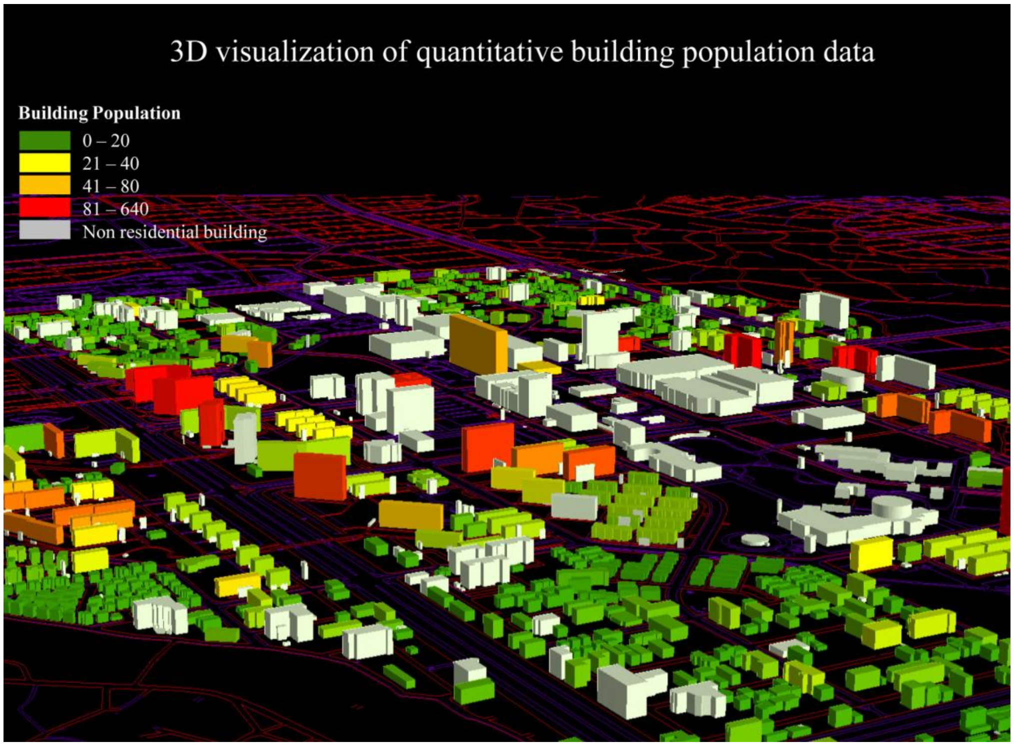


Figure 4. 3D visualization of quantitative building population data

The detailed study of "Estimation of Building Population from LIDAR-Generated DVM and Census Tracts" paper can be reached at following URL.
http://giswin.geo.tsukuba.ac.jp/sis/students/kokolwin/

Interactive Micro-spatial Population Analysis

Spatial analysis based on building population data is key benefit for disaster management teams in order to prepare humanitarian assistance when disaster occurred. They need specific quantitative amount of population with certain geographical unit such as 500m away from coastal lines in the case of tsunami strike or 5Km distance from earthquake's epicenter. We have developed a web-based interactive micro-spatial population analysis functions based on building population with other ancillary dataset such as public facility locations and transportation network. This web site can be reached at following URL.
http://land.geo.tsukuba.ac.jp/microspa/welcome.aspx

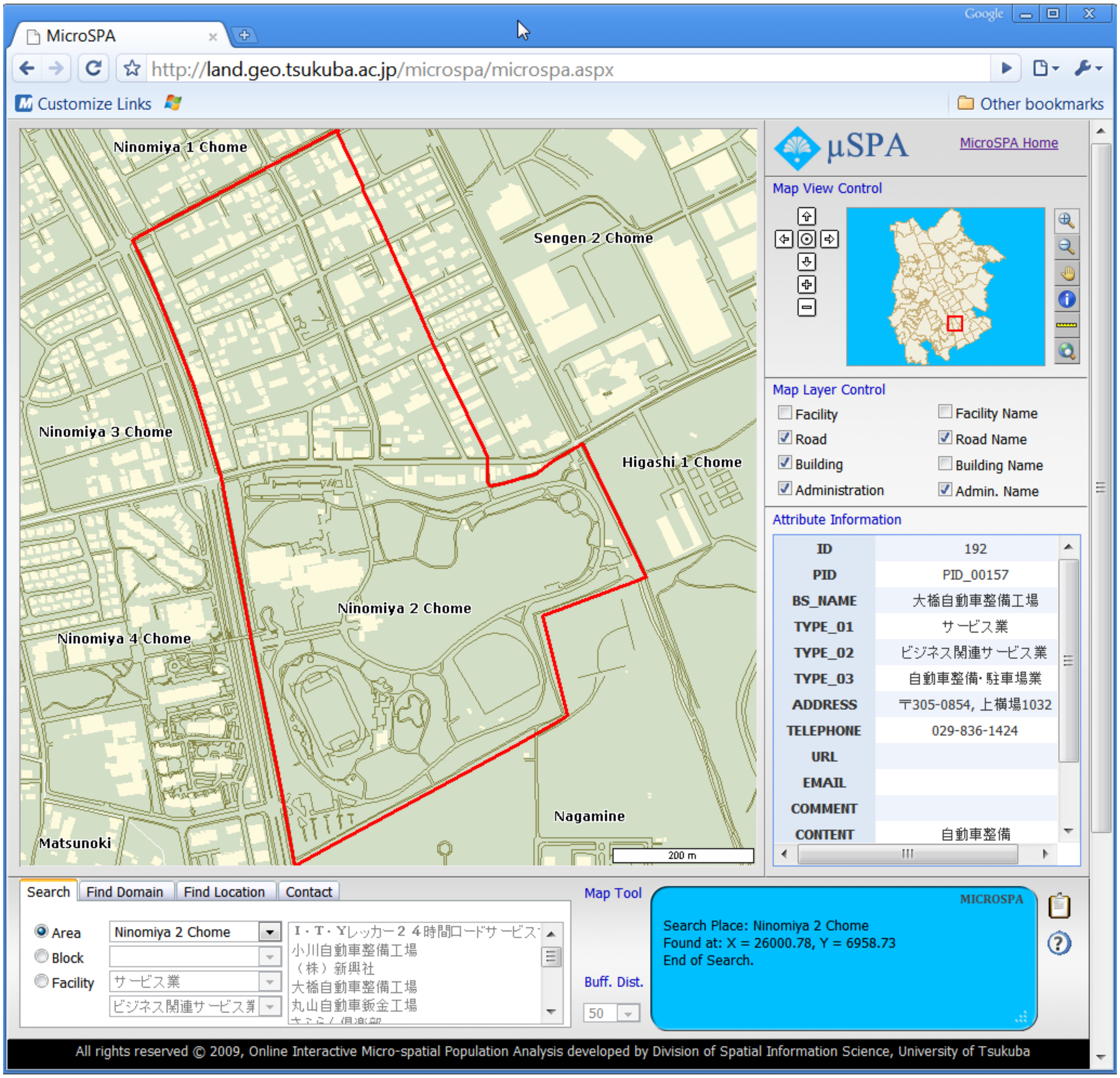


Figure 4. Graphical User Interface of online interactive micro-spatial population analysis

Map Layers and Data Sources

Following table 1 shows the list of map layers to be used in micro-spatial population analysis. Building footprints data was used from Zmap-TOWNII which is a commercial product from Zenrin Map Company and updated each year. Address of public facility and business locations data was acquired from NTT in text format. We geocoded (address matching) this NTT data using the geocoding engine developed by the Center for Spatial Information Sciences (CSIS), University of Tokyo and converted it into an ESRI Shape file in point format. Road centerline data was acquired from Geographical Survey Institute (GSI) in line features and then extracted real nodes and dangling nodes.

Table 1. List of map layers and data sources

Layer Order	Layer Name	Description	Feature	Visibility	Source
1	BUILDING	Building footprints with estimated population	Polygon	Visible	Zmap-TOWNII
2	FACILITY	Facility locations	Point	Visible	NTT Townpage
3	ROAD	Road outlines	Line	Visible	Zmap-TOWNII
4	ROAD_NODE	Road nodes	Point	Hidden	GSI
5	ADMIN-BND	Administration boundary	Polygon	Visible	Zmap-TOWNII

GIS Analytical Functions

Here we demonstrate such GIS analytical functions for local residents, business owners and urban planners to make their decision. These analytical functions can be divided into two main categories named as Find Spatial Domain and Find Location. Under Find Spatial Domain analysis category, users need to define their location and find the information such as finding the population mean center or facility mean center or connectivity mean center (road nodes) based on specific distance or buffered zone. To quantify the abstract idea of connectivity, Dill (2004) defines and makes the several measurement such as Link-Node Ratio, which is measured by dividing the number of links, (segments between nodes) in a study area by the number of nodes (intersections plus cul-de-sac termini); the Connected Node Ratio, which is a ratio of the number of street intersections to intersections plus the number of cul-de-sacs, thus capturing the number of connected nodes relative to the total number of nodes; and Intersection Density, which is simply the number of street intersections per unit of area. In This paper, we measure connectivity mean center which was calculated from road nodes coordinates (include all real and dangling nodes). Under the Find Location analysis category, users need to define the analysis boundary such as service zone or plan zone and find the suitable location inside the area.

Table 2. Summary of GIS Analytical Functions

Analytical Functions	**Find Spatial Domain**		**Find Location**	
Analysis Tool	**Circle**	**Poly Line**	**Polygon**	**Rect.**
Weighted building population mean center	☑	☒	☑	☑
Accessibility Index of building population	☑	☒	☒	☒
Facility mean center (selectable)	☑	☒	☑	☑
Accessibility Index of facility	☑	☒	☒	☒
Connectivity mean center (road nodes mean center)	☑	☒	☑	☑
Accessibility Index of connectivity	☑	☒	☒	☒
Total building population	☑	☑	☑	☑

☑ = Available ☒ = Not available

Measurement of Accessibility Indices

Measuring accessibility is a way of evaluating the availability and quality of basic services for spatial decision makers. As generally defined, accessibility reflects the ease of reaching needed or desired activities and thus reflects characteristics of both the land-use system (where activities are located) and the transportation system (how the locations of activities are linked) (Handy and Clifton 2001). There are several approaches to measure accessibility index or degree ranging from simple distance measurement to time and cost integrated measurement. Extensive academic literature on accessibility measures suggests many ways to define and measure accessibility, although examples of the actual use of accessibility measures in planning are relatively scarce (Handy and Clifton 2001). Accessibility is an important concept for urban planners because it reflects the possibilities for activities, such as working or shopping, available to residents of a neighborhood, a city, or a metropolitan area. Accessibility is determined by attributes of both the activity patterns and the transportation system in the area (Handy and Clifton 2001). Almost five thousands convenience stores are competing their locations in densely inhabited region like Tokyo (23 Wards; 621 km^2 with 8 million inhabitants (Okabe et al. 2002). Nowadays, interactive web based accessibility assessment and walkability score calculation is useful for home buyers and business owners.

The literature on accessibility measures has a long history. Most measures can be classified as one of three basic types (Handy and Niemeier 1997) named as Cumulative opportunities measures, Gravity-based measures and measure is based on random utility theory. In this paper, measurement of accessibility indices are based on three spatial elements; named as Building Population, Public Facility and Transportation Network. These three spatial elements are core components in urban system and interrelated with each other. For example, people need to find the ways to travel one place to another inside the city. They may need to find the shortest path to buy food and other services. On the other hand, business owners want to know their potential clients volume and how they can reach to their services. Urban planners or transportation planners need to know spatial distribution patterns of population and facilities in order to improve the urban transportation system inside the city. Each measurement is based on mean center of three spatial elements such as weighted building population mean center, weighted facility mean center and connectivity mean center (road nodes). Weighted building population mean center and weighted facility mean center were calculated by following equation.

$$\overline{X}w = \frac{\sum_i w_i.X_i}{\sum_i w_i}, \overline{Y}w = \frac{\sum_i w_i.Y_i}{\sum_i w_i} \tag{6}$$

Connectivity mean center was calculated by following equation.

$$\overline{\mathrm{X}} = \frac{\sum_{\mathrm{i}} \mathrm{X}_{\mathrm{i}}}{\mathrm{n}}, \overline{\mathrm{Y}} = \frac{\sum_{\mathrm{i}} \mathrm{Y}}{\mathrm{n}} \tag{7}$$

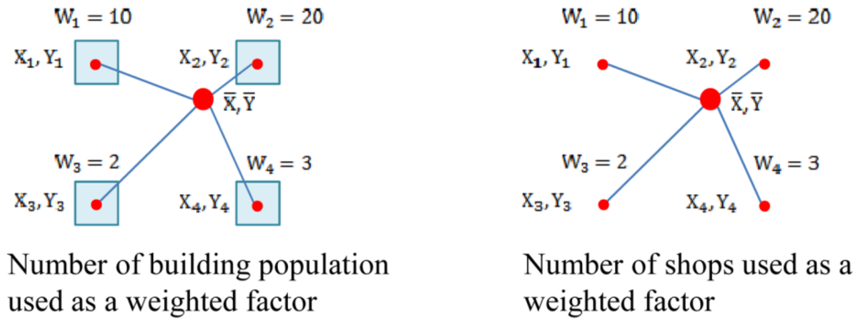


Figure 5. Calculation of weighted mean center

In order to measure accessibility index or degree, Euclidean distance was calculated between user's defined point and weighted mean center point. The accessibility index was computed by dividing Euclidean distance with circle radius. The circle radius is how far users are willing to travel or size of the service area in the case of business analysis. Calculation of facility mean center can be selected by specific business type such as noodle shop or convenient store. The index value is between 0 and 1. The interpretation of this value is depends on user's application. For example the smaller facility index is favorable for local residents and nearest competitor for business owners. Local residents can define various location and travel distance to evaluate their potential living place with facilities and connectivity indices. Business owners can assess their potential business location with population and connectivity indices. Urban planners can make future actions based on population and facility indices. Transportation planners can evaluate each bus station or railway station and their serving population by specific buffered distance.

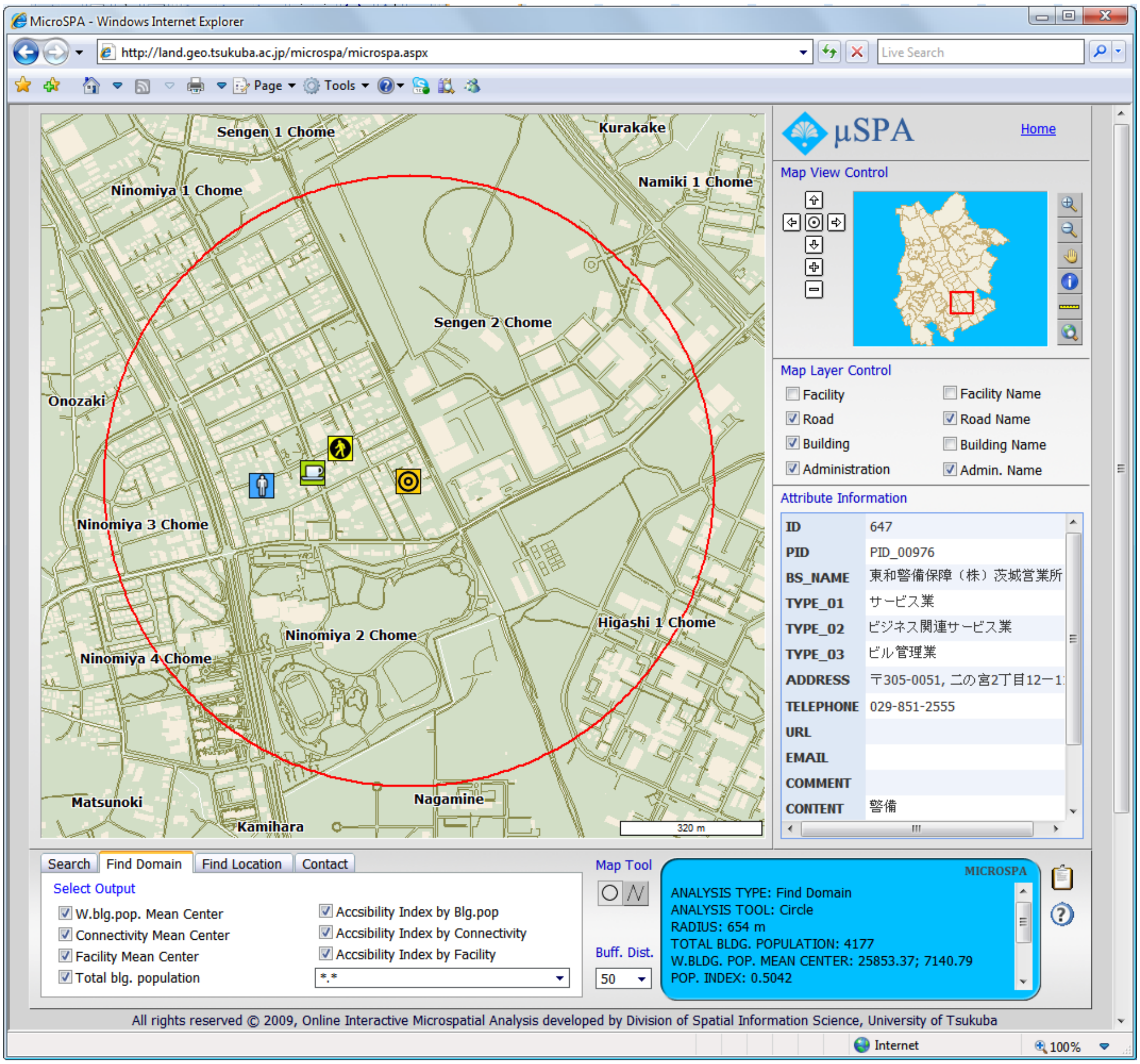
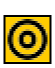


Figure 6. Find Spatial Domain analysis using Circle Tool to evaluate various locations

 Weighted building population mean center

 Weighted facility mean center

 Connectivity mean center (road nodes mean center)

 User defined point or location (center of the circle)

Under the line tool analysis, user can define buffer distance from the line and calculate total building population in the case of traffic noise impact studies or local community bus route planning (i.e., find shortest distance with highest population route).

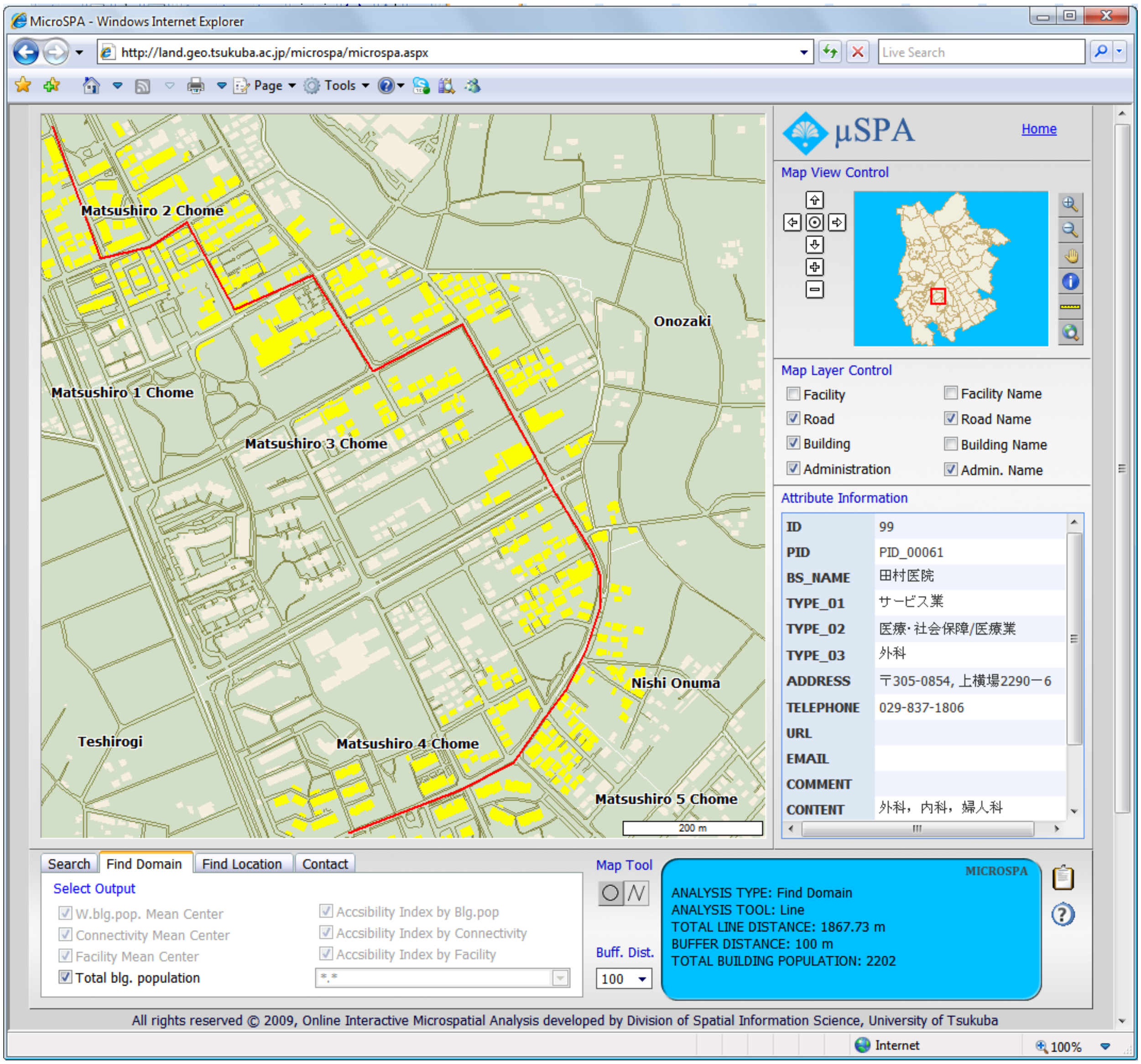


Figure 7. Find Spatial Domain analysis using Line Tool for local community bus route planning

Find location analysis module is based on user defined area by drawing polygon or rectangle. The calculation of building population mean center, weighted facility mean center and road nodes mean center are the same as find spatial domain analysis. The output results will be displayed as icons and their X, Y values in text format. The purpose of this module is to find potential local community center, business center, emergency response center and other public facility site sitting or estimate population from user defined zone (i.e., electoral zones or watersheds).

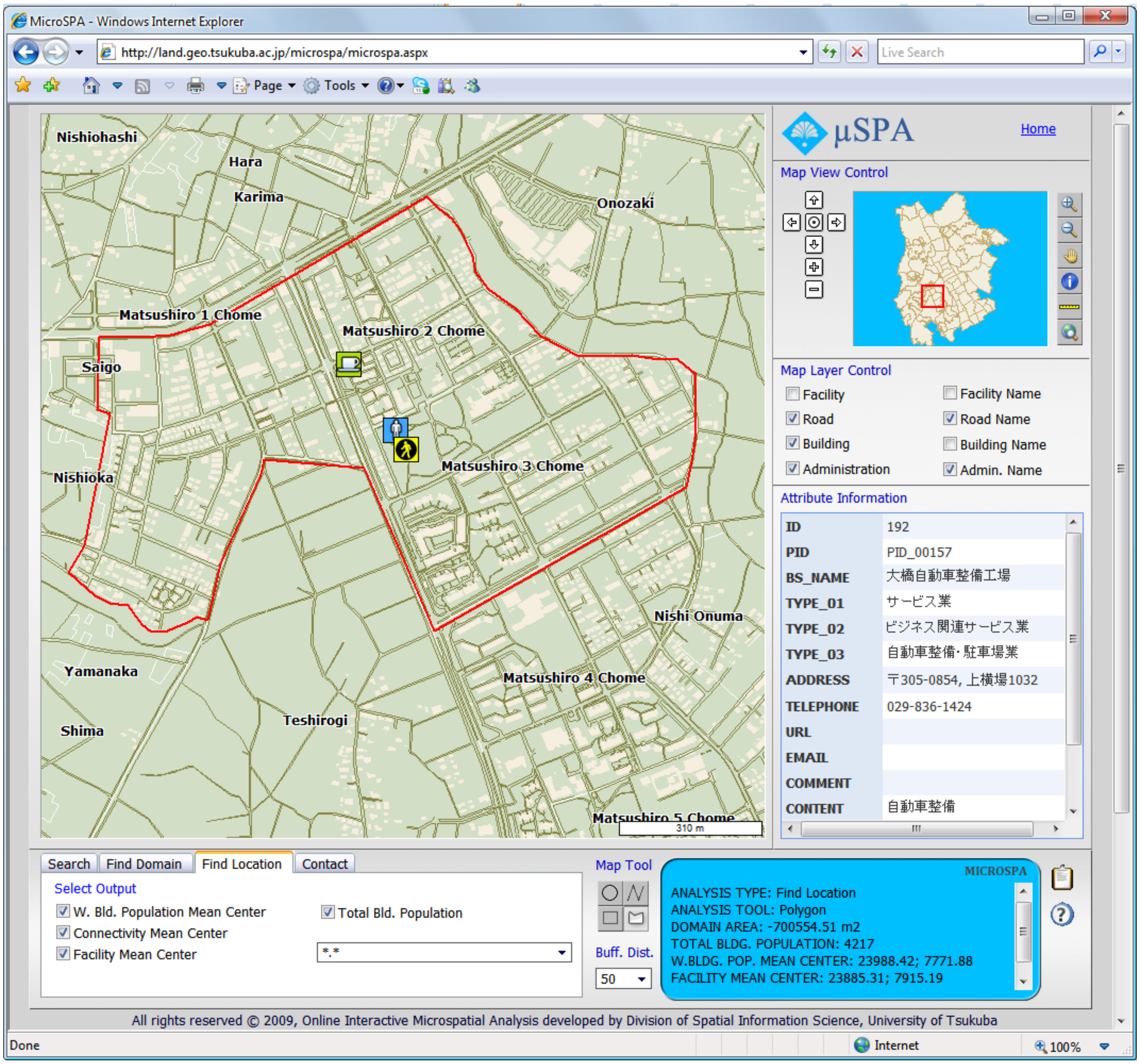


Figure 8. Find Location analysis using Polygon Tool for site selection

Development Platform

The overall system was built on Microsoft ASP.NET technology and VDS Technologies (Web Mapping Components for ASP.NET). ASP.NET is a web application framework marketed by Microsoft that programmers can use to build dynamic web sites, web applications and XML web services. AspMap for .NET from VDS Technologies is a set of high-performance, Web-mapping components and controls for embedding maps in ASP.NET applications (Web Forms).

CONCLUSION

Current population data represented in GIS are not suitable for use in micro-spatial population analysis. Spatial analysis functions performed inside the census tract do not acquire any significance changes in population. This paper introduced online interactive micro-spatial population analysis based on building population which was estimated from LIDAR derived DVM and number of floor information with census tracts for local residents, business owners and other urban planners. By integrating of LIDAR derived DVM with demographic data is key benefits for urban and city planners by improving their decision making processes in timely manner.

Acknowledgements

We would like to thank PASCO Corporation for providing the LIDAR data and Ortho images used in this study.

About the Authors

Ko Ko Lwin is currently enrolled as a PhD candidate at Division of Spatial Information Science, Graduate School of Life and Environmental Sciences, University of Tsukuba, Japan. He obtained his Master of Engineering degree from Center for Spatial Information Sciences (CSIS), Department of Civil Engineering, University of Tokyo, Japan.

Dr. Yuji Murayama is a Professor and department chair of Division of Spatial Information Science, Graduate School of Life and Environmental Sciences, University of Tsukuba, Japan. He teaches and researches issues on Human Geography, GIS and Spatial Analysis, and Urban Transportation Geography.

References

Bielecka, E., 2005, A dasymetric population density map of Poland. Proceedings of the International Cartographic Conference, July 9-15, A Coruña, Spain, CD Rom

Dill, Jennifer, Measuring Connectivity for Bicycling and Walking In TRB 2004 Annual Meeting, Transportation Research Board, National Research Council, Washington, D.C., 2004, CD Rom

Haala, N., Brenner, C., 1997, Generation of 3-D City Models from Airborne Laser Scanning Data. EARSel Workshop on LIDAR Remote Sensing of Land and Sea, Tallinn, Estonia.

Handy, S.L. and D.A. Niemeier. 1997, Measuring Accessibility: An Exploration of Issues and Alternatives, Environment and Planning A 29:1175–94.

Handy, S. L. and K. J. Clifton, Evaluating Neighborhood Accessibility: Possibilities and Practicalities. Journal of Transportation and Statistics, September/December, 2001.

Hug, C., 1997, Extracting Artificial Surface Objects from Airborne Laser Scanner Data. Grün, Baltsavias, Henricsson (Eds.), Ascona: Automatic Extraction of Man-Made Objects from Aerial and Space Images II (Basel Birkhäuser Verlag).

Kim, H., Kim, Y., Cho, W., Seo, J., 2000, The Fusion of Photogrammetric Image Laser and Scanning Data. Proceedings of the American Society for Photogrammetry and Remote Sensing Conference, Washington, D.C.

Mrozinski, R.D., Jr., Cromley, R.G., 1999, Singly and doubly constrained methods of areal interpolation for vector-based GIS. Transactions in GIS 3: 285-301.

Nichol, J.E., 1996, High-Resolution Surface Temperature Patterns Related to Urban Morphology in a Tropical City: A Satellite-Based Study. J. Appl. Meteorol. 35(1):135-146.

Okabe, A., Okunuki, K., Funamoto, S., Okano, K. and Itoh, F. (2002) "Software for spatial analysis on a network", Proceedings of Specialist Meeting on Spatial Data Analysis Software Tools, Santa Barbara, May 10-11, 2002, CD Rom

Books

Belward, A.S. and Valenzuela, C.R. (Editors) 1991 Remote sensing and geographic information systems for resource management in developing countries, Kluwer Academic Publishers, Boston, MA, 1991.

Mitchell, A. 2005, The ESRI Guide to GIS Analysis: Volume 2: Spatial Measurements and Statistics, ESRI Press, Redlands, California.

Openshaw, S. 1992, Learning to Live with Errors in Spatial Databases. In Goodchild, M and S. Gopal (eds.) The Accuracy of Spatial Databases. New York: Taylor and Francis 263-276.

Internet

Sleeter, R., 2008, A new method for mapping population distribution: U.S. Geological Survey Fact Sheet 2008-3010, 2 p. http://pubs.usgs.gov/fs/2008/3010/