Tsukuba Geoenvironmental Sciences, Vol. 8, pp. 9-15, Dec. 28, 2012

Mapping the housing types from LIDAR data for micro-scale spatial analysis: A case of Tsukuba City

Ko Ko LWIN* and Yuji MURAYAMA*

Abstract

Spatial distribution patterns of housing types are required for public facility management, disaster and emergency preparedness, retail market analysis and other demographic studies. Because housing type is strongly related to household unit (i.e. Single or Family units). In this paper, we identify the spatial distribution patterns of housing types (i.e. Single Multiple Unit (SMU), Family Multiple Unit (FMU) and Family Single Unit (FSU), based on building footprints and their shapes and sizes, identified from Light Detection and Ranging (LIDAR) data. We also utilize an Internet telephone directory NTT iTownPage data to separate residential and non-residential buildings. The result was validated by number of household units by administrative block acquired from the city office. We found that the housing types are well associated with household units per administrative block known as Chome (smallest administration unit). This result can be used in micro-scale spatial analysis for retail market analysis, and public facility management where detailed census data is not available.

Keywords: Spatial distribution patterns, housing types, LIDAR and NTT iTwonPage

1. Introduction

LIDAR data is used widely to identify three-dimensional landscape features in many real-world applications, such as radio and microwave signal analysis for communication site planning, three-dimensional landscape visualization for urban planning, geology and mining planning, such as estimation of ore deposit or excavated volume, and hydrological modeling. Recently, the application of LIDAR data in human geography is increasing in terms of estimation of building population, based on its surface area and volume (Lwin and Murayama, 2009), and the detailed population distribution map, known as a dasymetric map, can be generated by integrating with LIDAR data (Lwin and Murayama, 2010).

On the other hand, building size and shape are related to housing types and utilization of LIDAR data to identify housing types is necessary for public facility management and retail market analysis. Housing types can be grouped into Single Unit which includes apartments where single people are living and Family Unit which includes single houses and mansions where family people are living. For example, single-unit places require more convenience stores and fast food shops, while family unit places require more grocery shops, schools and parks. Other public facilities are also needed for both household types, such as gyms, movie theaters and home appliance shops.

2. Methodology

2.1. Study Area

The study area is part of Tsukuba City, a planned city for academic and scientific purposes and the home of the University of Tsukuba. A total population of 39,094 (2006), located in 29 small administrative blocks, was used for this study. Figure 1 shows the study area, population, number of households, and household percentage by administrative block. Household percentage for each administrative block is the household units divided by population, multiplied by 100. More household percentage means more single people (*i.e.*, students, part-time workers, etc.) are living in this administrative block.

In this study, we used LIDAR data in ESRI point format for both a Digital Surface Model (DSM) and a Digital Terrain Model (DTM), provided by PASCO Corporation, whose vertical RMSE is +15 cm. Each DSM and DTM scene is 800 m wide and 600 m long. There were 45 DSM and 45 DTM scenes used for the whole study area. DSM points are, on average, 0.9 m apart, and DTM points are at a regular 5 m spacing. PASCO also provided 8 cm ortho images, which were acquired along with the LIDAR survey.

2.2. Data Processing

2.2.1. Digital Height Model (DHM) and Digital Volume Model (DVM) Generation

Traditionally, stereo images matching is a standard photogrammetric technique used to generate a Digital Surface Model (DSM). However, this technique is only good for an open, smooth terrain surface. The quality of Digital Surface Model (DSM) in built-up areas is poor, owing to occlusions and height discontinuities (Haala and Brenner, 1999). LI-

* Faculty of Life and Environmental Sciences, University of Tsukuba

Table 1. List of geospatial data used in this study

Data and Source	Description	Purpose
Digital Surface Model (DSM) (Source: PASCO Corp.)	Point feature in ESRI shape format Average point spacing is 0.9 m One scene (800 m X 600 m) Acquired in 2006 A total of 45 scenes were used	To compute Digital Height Model (DHM) where: DHM = DSM – DTM
Digital Terrain Model (DTM) (Source: PASCO Corp.)	Point feature in ESRI shape format Each point is 5 m regular spacing One scene (800 m × 600 m) Acquired in 2006 A total of 45 scenes were used	To compute Digital Height Model (DHM) where: DHM = DSM – DTM
Orthoimages, (Source: PASCO Corp.)	GeoTIFF format 8 cm × 8 cm spatial resolution RGB True Color One scene (800 m × 600 m) Acquired in 2006 (along with LIDAR survey) A total of 45 scenes were used	Used as a base map and landscape visualization
Building Footprints (Source: ZMapTOWN-II product by Zenrin Co. Ltd.)	Polygon feature in ESRI shape format, which includes building name, block number, number of floors, and other attribute information Acquired in 2006 Map Scale 1:2,500	To identify housing types
Administrative blocks (Source: ZMapTOWN-II product by Zenrin Co. Ltd.)	Polygon feature in ESRI shape format, which includes administration boundary units and names Acquired in 2006 Map Scale 1:2,500	To validate the result

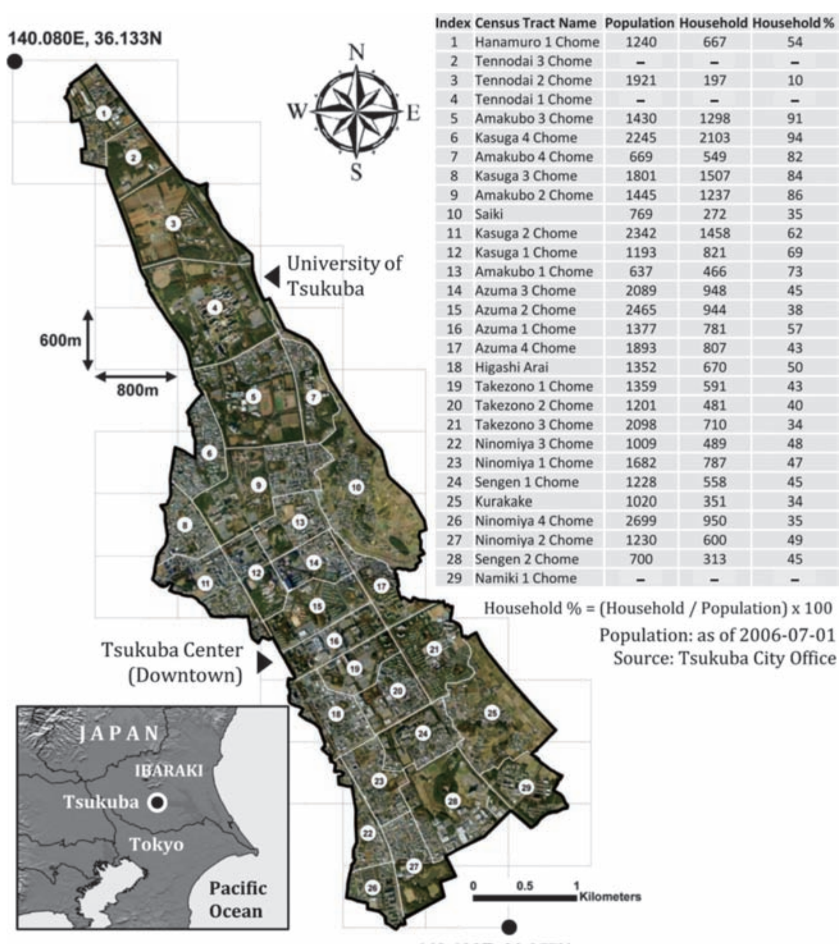


Fig. 1 Study Area

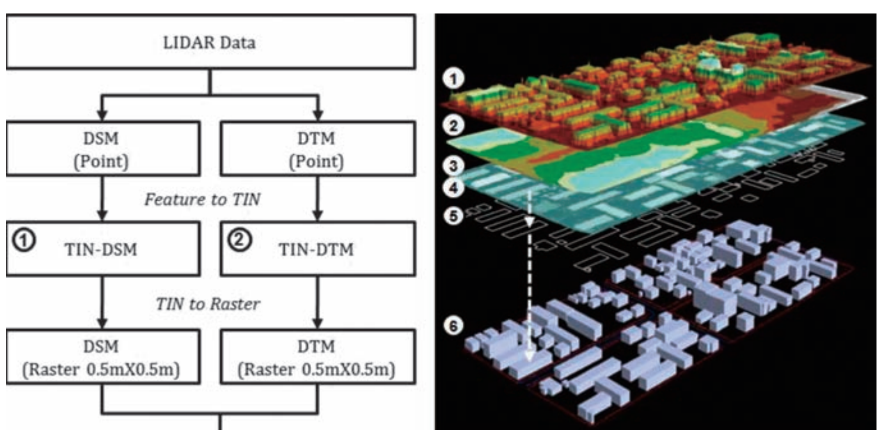
DAR techniques have been studied and utilized since the early 1960s, but appear to have become more prominent in the past few years. We have found LIDAR applications in a wide variety of fields of study, including atmospheric science, bathymetric data collection, law enforcement, telecommu-nications, and even steel production (Maune *et al*., 2000). Since LIDAR is operated with much shorter wavelengths, it has higher accuracy and resolution than microwave radar (Jelalian, 1992).

Choosing the appropriate surface generating method for DSM and DTM is important in LIDAR data processing, since surface height information is collected as points. Figure 2 shows the detailed procedure of generating DHM and DVM surface models. In this process, both DSM and DTM point features were converted into a Triangulated Irregular Network (TIN) model (*i.e.* TIN-DSM and TIN-DTM). Under the ArcGIS program, the TIN process allows users to convert multiple scenes at one time. This reduces the time for mosaicking. Moreover, the TIN process is faster than other interpolation processes such as Inverse Distance Weighed (IDW), Spline and Kirging. Each TIN-DSM and TIN-DTM was then converted into a raster format, setting the spatial resolution to 0.5 m. We subtracted the DTM from the DSM raster layers to achieve the Digital Height Model (DHM). This DHM raster layer was multiplied by the cell surface area (*i.e.* 0.25 m2) to convert to a Digital Volume Model (DVM).

2.2.2. Spatial Adjustment

In this study, we used the building footprint data (ZMap-TOWNII) from Zenrin, in a shape file polygon format. When the footprint data was overlaid with the LIDAR-derived DHM, it became apparent that there were some spatial adjustments needed. To improve the visualization, and to remove the noise from DHM data, we applied a 5x5 kernel low-pass filtering process. We also filtered out pixels with heights < 2 m, and performed hill shading. This removed the cars, small bushes and other objects with heights less than 2 m.

To perform spatial adjustment, we overlaid building footprints on the shaded DHM, and adjusted the building footprint based on the shaded DHM. As a digital form, hillshade is a grid that encodes the reflectance value off an elevation surface, given a light source at a certain theoretical position in the “sky”. The hillshade function obtains the hypothetical illumination of a surface by determining illumination values for each cell in a raster. It does this by setting a position for a hypothetical light source, and calculating the illumination values of each cell in relation to neighboring cells. It can greatly enhance the visualization of a surface for analysis, or graphical display. By applying different azimuth angles, we can obtain different side



PID	AREA	H_MEAN	H_SOURCE	H_STD	V_SUM	V_SOURCE
PID_720	174.5978	5.5873	LIDAR	0.9845	972.1980	LIDAR
PID_721	176.9578	6.0804	LIDAR	0.5258	1076.2400	LIDAR
PID_722	188.9076	6.4752	LIDAR	1.5185	1222.2000	LIDAR
PID_724	82.1189	3.9560	LIDAR	1.8953	325.3770	LIDAR
PID_725	43.9445	2.4470	LIDAR	0.5332	107.0560	LIDAR
PID_726	74.8091	2.9327	LIDAR	0.6126	219.2170	LIDAR
PID_727	144.4233	3.2186	LIDAR	1.2545	466.7000	LIDAR
PID_728	17.2848	3.7575	LIDAR	0.8217	64.8162	LIDAR
PID_729	1134.1210	8.4507	LIDAR	4.7853	9585.2002	LIDAR
PID_730	247.4970	5.3646	LIDAR	1.1407	1327.7400	LIDAR
PID_732	144.5287	4.4927	LIDAR	1.8652	650.3230	LIDAR
PID_733	202.2426	6.1844	LIDAR	1.6201	1252.3500	LIDAR
PID_734	125.2634	6.0588	LIDAR	1.5820	761.8950	LIDAR
PID_735	12.0599	2.4802	LIDAR	0.3349	29.7623	LIDAR
PID_736	439.6767	10.4085	LIDAR	3.6050	4579.7500	LIDAR
PID_737	216.5759	5.7869	LIDAR	1.8010	1251.4100	LIDAR

Fig. 2 LIDAR data Processing Work Flow (Source: Lwin and Murayama, 2010)

Table 2. List of non-geospatial data used in this study

Data and Source	Description	Purpose
Tsukuba City resident registration (Source: Tsukuba City Office)	Microsoft Excel Sheet Male, female, household units and total population by administrative block	To compute building population
	To join with administrative unit polygons from ZMapTOWN-II data, and use as administrative block population	
	Updated monthly	
	Used 2006-07-01 resident registration information	
iTownpage (Source: Nippon Telegraph & Telephone Corp. (NTT))	Comma Separated Value (CSV) format	To separate residential and non-residential buildings
	Information of public facilities and business locations	
	Business name, business category, business subcategory, business contents, address, phone, URL, etc.	To detect mixed building use type

views of the building shadow edges (see Figure 3).

2.2.3. Building Height and Roof Standard Deviation Extraction

After spatial adjustment was performed, the building height and volume were extracted from DHM and DVM, respectively, by applying Zonal Statistics as Table in ArcGIS (ToolBox > Spatial Analysis Tools > Zonal > Zonal Statistics as Table). This function allows users to summarize the values of a raster (*i.e.* DHM and DVM raster) within the zones of another dataset (i.e. building footprint polygons), and reports the results to a table. Later, this table was joined with building footprint polygons, based on the same polygon ID. We extracted average and standard deviation (SD) of building height from DHM, which was used to identify smoothness of the building roofs (*i.e.* low SD value gives flat roof and high SD value gives irregular roof form; Figure 4).

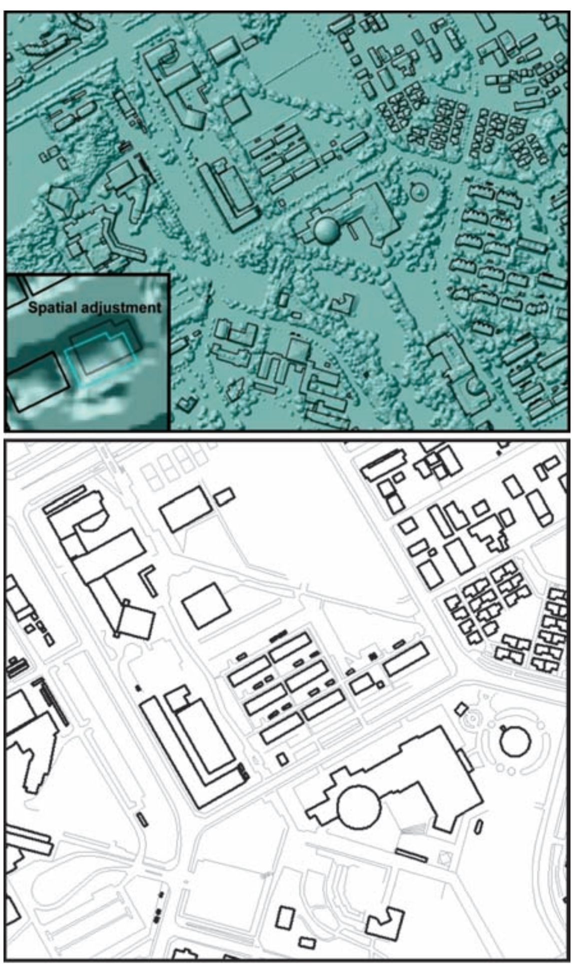


Fig. 3 Spatial adjustment performing between 2m above shaded Digital Height Model and building footprints

2.2.4. Conversion of iTownpage CSV into Point Feature

In recent years, the emergence of spatial information technologies has allowed us to transform non-spatial information into spatial datasets, such as geo-coding and address-matching techniques, that enable us to convert address information into x, y coordinates. In this study, we downloaded iTownpage data from the Nippon Telegraph & Telephone Corp. (NTT) website, which includes business name, category, subcategory, content, address, and other information in a Comma Separated Value (CSV) format. For separating residential buildings from the non-residen-

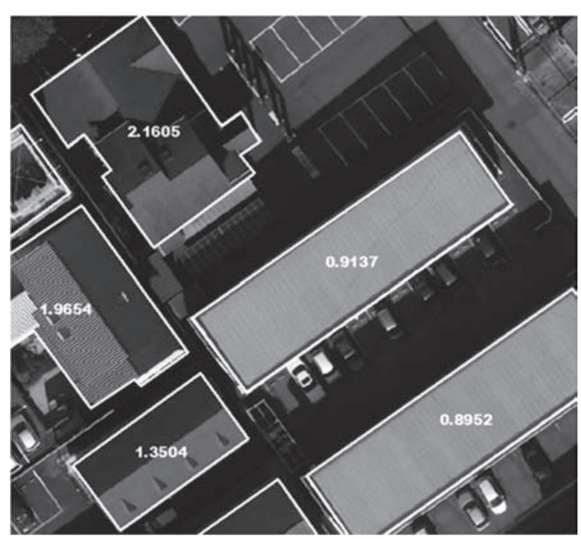


Fig. 4 Average standard deviation of building heights on 8 cm Orthoimage

tial, this CSV data was converted into ESRI point features, using commercial geo-coding software with building level accuracy.

2.2.5. Separation of Residential vs. Non-Residential Building

Step 1: Filter by empty population administrative blocks.

Remove all building footprints that fall inside the empty population administrative bocks such as university and research center campus.

Step 2: Filter by Building Area and Height.

Remove all building footprints with areas of less than 20 m^2 and with a height of less than 2 m, thus removing bicycle stand roofs, garbage boxes, porticos and other footprints not occupied by people.

Step 3: Filter by Public Facilities.

In this step, we used the facility point layer, which was generated from NTT iTownpage, to identify large-scale public facilities, such as financial institutions, governmental organizations, educational organizations, etc. This can be done by intersecting the NTT Facility point data with the building footprints polygons. Although the geo-coding accuracy of NTT data conversion is building level, we applied a 3 m tolerance, buffered to each facility point during the intersecting process.

Step 4: Manual Filtering

This step involved the removal of other unpopulated building footprints, which cannot be detected automatically, such as large building footprints for multi-story car parking lots, and some farming units in rural areas. In this process, we used 8 cm orthoimages to interpret the landscape features visually.

3. Results and Discussion

3.1. Identification of Housing Types

In our study area, the building height and shape are normally related to a family unit. These buildings can be grouped into three housing types; for example: Single Multiple Unit (SMU); Family Multiple Unit (FMU); and Family Single Unit (FSU). SMUs (Figure 5A) are designed especially for college students, factory workers, and other part-time workers. Usually, this type of building is not more than three or four floors, and the shape of the building is mostly cubic form (*i.e.* a regular shape to reduce the construction cost and rental price), and the standard deviation of their heights is very low. This kind of building is found commonly in the surroundings of Tsukuba University campus. FMUs (Figure 5B) are more than three floors, with either regular or irregular form, with their height standard deviation being either high or low. Sometimes, additional sections are attached to the main building, such as an elevator tower, swimming pool or multi-story car parking lot. FSUs (Figure 5C) are standalone buildings whose shape is almost irregular, and standard deviation of their heights is always high. Normally, this kind of building has one or two floors, and is mixed with paddy fields. Table 3 shows the classification categories of three housing types in the study area.

3.2. Housing Types Mapping

Figures 6–8 show the spatial distribution patterns of housing types in various administration blocks in the study area. Many SMUs are found near the university campus, and many FMUs are found in central area, while the many FSUs are found outside of the central area, where paddy fields are abundant.

3.3. Results Validation

We validated our results with the percentage of household units, calculated from administrative unit data provided by the city office. Figure 9 compares the percentage of single multiple units (*i.e.* Total number of SMUs / Sum of FMUs + FSUs) with household percentages from administrative unit data. Many SMUs are found in area surrounding the university campus, where student apartments are abundant. Additional public facility data, such as daycare centers, schools and office locations, are required for

Fig. 5 Common housing types and building shapes in study area. (A) Single Multiple Unit (SMU), (B) Family Multiple Unit (FMU), and (C) Family Single Unit (FSU) (high height standard deviation), from side view and aerial view

Fig. 6 SMU housing type abundant areas (near the university campus)

Fig. 8 FSU housing type abundant areas (away from the central area)

Fig. 7 FMU housing type abundant areas (near the central area)

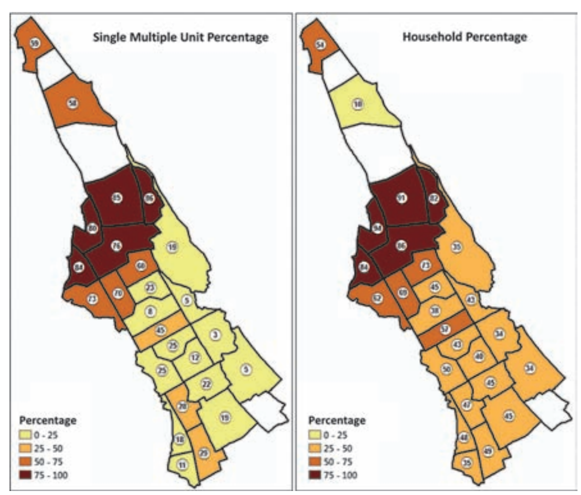


Fig. 9 Comparison of SMU housing type percentage and household percentage

Table 3. Housing type classification categories

	Single Multiple Unit	**Family Multiple Unit**	**Family Single Unit**
Height	Medium (0–15 m)	High (> 15 m)	Low (< 10 m)
Surface Area	Medium	Large	Small
Roof Shape	Low (< 2 SD)	Medium (2–4 SD)	High (> 4 SD)

FMU and FSU validation by performing proximity analysis, since many family members prefer to live near school and office facilities.

4. Conclusion

Although LIDAR data and building footprint data are considered as high level geospatial data in terms of cost, and high technology, with sophisticated instruments for data acquisition and require insensitive data processing steps, the identification of housing types by means of LIDAR data is very promising due to its finer spatial resolution and accurate height information. Moreover, the building footprint level of geospatial data is important for micro-scale spatial analysis, such as the identification of housing types by specific buffer distance, along the road or user-defined point.

Acknowledgement

We would like to thank the PASCO Corporation for providing LIDAR data and 8cm Orthoimage for this study. This research is supported financially by the Grant-in-Aid for Scientific Research A, JSPS, 2012-05, Application strategies of advanced spatial analysis techniques in the fields of geographic information science and urban engineering (Chief Investigator: Tsutomu Suzuki, University of Tsukuba).

References

Haala, N. and C. Brenner, 1999. Extraction of buildings and trees in urban environments, *ISPRS Journal of Photogrammetry and Remote Sensing*, **54**(2-3): 130-137.

Jelalian, A., 1992. Laser Radar Systems, Artech House, Boston.

Lwin, K. K. and Y. Murayama, 2010. Development of GIS tool for dasymetric mapping, *International Journal of Geoinformatics*, **6**(1):11-18.

Lwin, K.K. and Y. Murayama, 2009. A GIS approach to estimation of building population for micro-spatial analysis, *Transactions in GIS*, **13**(4): 401-414.

Maune, D., C. Daniel and J. Damron, 2000. LIDAR and IFSAR: Pitfalls and opportunities for our future, *Proceedings of the American Society for Photogrammetry and Remote Sensing Conference*, Washington, DC.

Received 22 August 2012
Accepted 24 October 2012