X165-010 **Room: 101A** **Time: May 28 16:11-16:24**

Development of 3D Tenant Monitoring dataset in Broad Urban Area by Spatio-temporal Integrating Digital Maps and Yellow Page Data

Yuki Akiyama[1]; Takeshi Shibuki[2]; Ryosuke Shibasaki[3]

[1] Frontier Science, The University of Tokyo; [2] Civil, Univ. of Tokyo; [3] CSIS,UT

http://shiba.iis.u-tokyo.ac.jp/

So far many kinds of dataset were developed for urban analysis. However, it has been difficult to create dataset with high spatial accuracy and reliability in broad area. Detailed time-series dataset are demanded to monitor diverse changes in urban space. Many studies also pointed out the need of such dataset. Thus, the development of the dataset all over Japan is most rewording for urban analysis of Japan.

We focus on changes of tenants for monitoring urban space. The tenant means all kinds of shops and offices without personal residences. Tenant data is collected from digital house maps and yellow page. All tenants in the dataset have information of time-series variation, namely, continuations, changes, emergence and demise. In addition, almost all tenants have detailed location information including building information and business categories.

The dataset is developed as follows flow. First, house maps and yellow page are integrated, because only yellow page has business categories. Secondly, multi-year digital house maps which have yellow page data are integrated.

Thus, a method to integrate different kinds of data is required. Moreover, a method to specify time-series variation of tenants between different two years is required.

To realize these methods, development of spatial integration method based on longitude, latitude, address and building information and evaluation method of tenant names are main tasks.

Not all tenant locations have complete location data, for example, loss of address and building information. Therefore, one tenant is integrated a tenant which has most similar location.

The location integration method is able to deal with not only longitude and latitude, but also nonquantifiable location information, namely, address and building information. It differs greatly from typical GIS software.

An identification of tenants which are integrated is judged based on their tenant names. Since tenant names are spelled in different ways even same tenants, a quantification method of name similarity is needed to realize this identification.

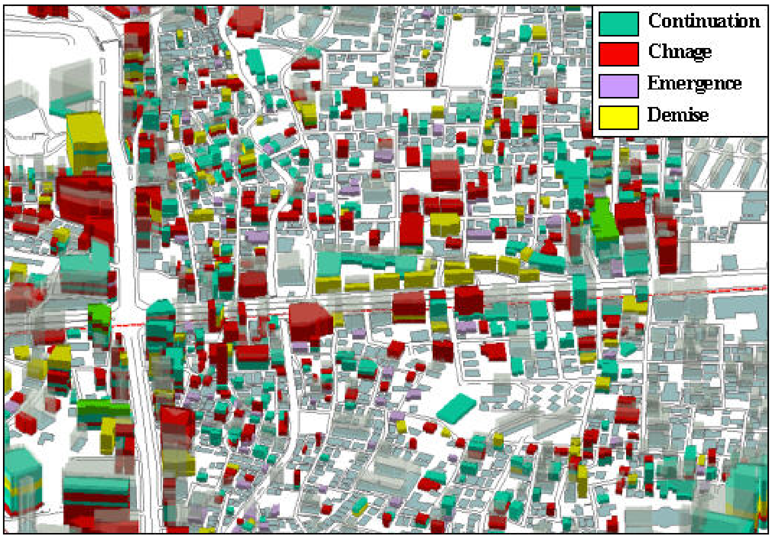
Tenant names are identified using the n-gram. The n-gram is able to quantify a similarity between two words or texts. However, accurate name identification is difficult using only the n-gram by the effect of noisy words, such as frequent words and geographic names. These words are removed using libraries of frequent words. In addition, it is possible that English texts are translated into Japanese.

We compared system results with manual results to assess the processing accuracy. 500 samples are extracted form the special words of Tokyo in random order. As a result, 94.00% integration result of each data, 95.88% of multi-year yellow page and 93.00% of multi-year house maps are accurately integrated.

To assess the reliability of source data we conducted field surveys around Akabane station and Shimo-takaido station, next, compared tenant locations and names of source data with results of field surveys. As a result, 88.16% of house maps and 87.74% of yellow page data accord real space in Akabane, and 89.51% of house maps and 91.51% of yellow page data accord in Shimo-Takaido.

Finally, we introduce an example of results. Figure shows 3D maps of tenant variations between 2000 and 2005. Such 3D spatio-temporal data can be created all over Japan.

The spatio-temporal dataset which can monitor on tenant scale has been developed. The spatial integration method based on not only longitude and latitude but also nonquantifiable information and the identification method of texts which considers ambiguity of texts have been developed. Moreover, the processing of data development has been automated. In conclusion, it is the practical method that variation in broad area can be continuously monitored at low cost.



Continuation
Chnage
Emergence
Demise