International Journal of Earth, Energy and Environmental Sciences
ISSN: 2517-942X
Vol:4, No:6, 2010

Automatic Detection and Spatio-temporal Analysis of Commercial Accumulations Using Digital Yellow Page Data

Yuki. Akiyama, Hiroaki. Sengoku, and Ryosuke. Shibasaki

***Abstract*—**In this study, the locations and areas of commercial accumulations were detected by using digital yellow page data. An original buffering method that can accurately create polygons of commercial accumulations is proposed in this paper.; by using this method, distribution of commercial accumulations can be easily created and monitored over a wide area. The locations, areas, and time-series changes of commercial accumulations in the South Kanto region can be monitored by integrating polygons of commercial accumulations with the time-series data of digital yellow page data. The circumstances of commercial accumulations were shown to vary according to areas, that is, highly- urbanized regions such as the city center of Tokyo and prefectural capitals, suburban areas near large cities, and suburban and rural areas.

***Keywords*—**Commercial accumulations, Spatio-temporal analysis, Urban monitoring, Yellow page data

I. Introduction

MANY shopping areas follow a course of decline because of vacant stores, aging shopping facilities, changes in consumer needs, and a competitive environment. Some rural and suburban shopping areas have crumbled in terms of the continuity of their shopping streets and diversity of business categories due to an increase in the number of closed stores. As a result, many shopping areas have lost the convenience and attraction they offered consumers in the past by collecting stores in a single area;, these shopping areas are now declining at an alarming pace. On the other hand, some shopping areas are actively attracting new shops and offices. Therefore, a method to monitor the actual state of shopping areas is needed.

However, monitoring the actual state of shopping areas in local cities, especially small rural towns, is difficult because there is little existing spatial data in such areas.

Many studies have tried to monitor the actual state of commercial accumulations. For example, Ato et al. [1] revealed the time-series changes and its problems in the shopping areas in front of train stations in the suburbs of the Tokyo metropolitan area: they developed distribution data of emerging and declining stores. They acquired detailed insights into shopping areas through field survey and bibliographical surveys. However, methods used in many previous studies including theirs require large amounts of labor and time, ; in addition they lack versatility because many previous studies are case studies for limited areas.

The objective of this paper is to propose an automatic detectable method for shopping areas and commercial accumulations using the digital yellow page data of Japan. The area under study was the South Kanto region which includes the special wards of Tokyo (the core and the most populous part of Tokyo).

In addition, changes in scale (number of tenants), changing rates of tenants, and ratios of chain store in each commercial accumulation were examined. The time-series changing data for each tenant between different years were required to reveal the changing rates of tenants in each commercial accumulation. These data were realized by the Spatio-temporal data integration system introduced in the next section.

II. Spatio-temporal data integration system

In this study, a system was developed that can integrate tow different datasets with name information, e.g., shops or company names and addresses based on location information. The system integrates two different sets of data based on location information (address, longitude, and latitude, and building information). At the same time, the system recognizes the identity of each tenant’s names. The system realizes 3D spatial integration, the cleaning of frequently appearing words in name information, and identity verification of names with name fluctuations. The system can also integrate and identify not only different kinds of dataset with name and address information, but also old and new datasets. Continuation, change, emergence and closure of all tenants can be monitored.

The dataset shown in “Fig. 1” monitors the tenant changes for each floor and room. The dataset is realized to integrate time-series yellow page data between 2000 and 2005; it was developed by the proposed system using digital house map (Zmap TOWN II; ZENRIN CO., LTD.). In addition, the macro scale dataset showed in “Fig. 2” can be developed.

Yuki Akiyama is with Graduate School of Frontier Science, The University of Tokyo, Institute of Industrial Science, Cw-503 4-6-1, Komaba, Meguro-ku, Tokyo 153-8505, Japan (phone: +81-3-5452-6417; fax: +81-3-5452-6414; e-mail: aki@iis.u-tokyo.ac.jp).

Hiroaki Sengoku is Graduate School of Frontier Science, The University of Tokyo, Cw-503 4-6-1, Komaba, Meguro-ku, Tokyo 153-8505, Japan (e-mail: xianshi@csis.u-tokyo.ac.jp).

Ryosuke Shibasaki is with Center for Spatial Information Science, The University of Tokyo, 5-1-5, Kashiwanoha, Kashiwa City, Chiba 277-8568, Japan (e-mail: shiba@csis.u-tokyo.ac.jp).

In this study, spatio-temporal yellow page data obtained between 2000 and 2005 were used to develop this system. The digital yellow page data were from the *"Town Page Database"* published by NTT Business Information Service, Inc.

Details on the proposed system and method and the reliability of the time-series yellow page data used in this paper are available in previous papers (Akiyama et al.[2][3]).

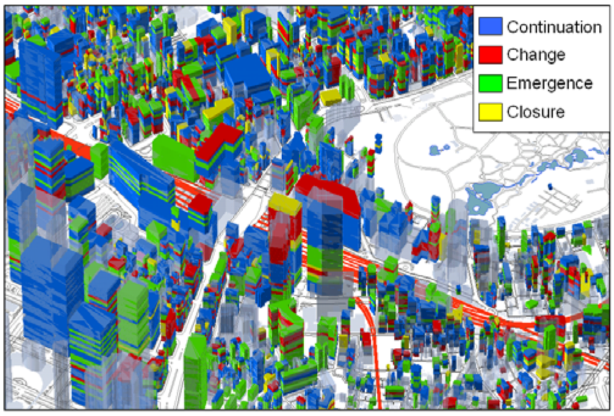


Fig.1 3D map of tenant changes in each floor between 2000 and 2005 around Shinjuku terminal (busiest train station in the world in terms of number of passengers)

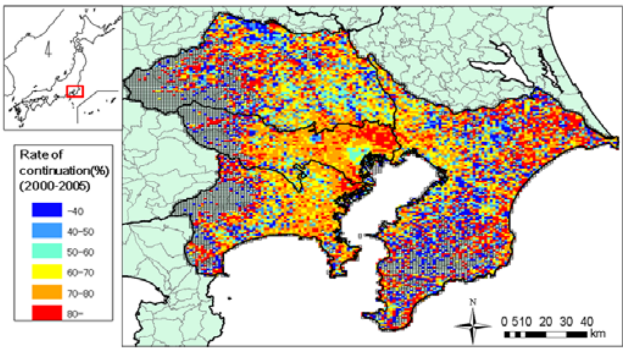


Fig. 2 Continuation rate of tenants between 2000 and 2005 in South Kanto region (1km grid map)

III. DEVELOPMENT

A. Definition of commercial accumulations

Commercial accumulations ("CAs") are defined in this paper as shopping districts and streets, such as shopping areas and streets around train stations, shopping districts in city core areas, and shop clusters in sightseeing areas. These commercial districts are called *"shotengai"* in Japanese. "Fig. 3" shows an example of a bustling *shotengai* street. On the other hand, there are many declining *shotengai* districts like that shown in "Fig. 4", especially in smaller cities.

Roadside shops along major roads and large shopping malls with their tenants that emerge in recent years are also included under this definition. In this study, a method that can detect CA distributions is required.

Fig. 3 Bustling *shotengai* (Togoshiginza Syotengai, Tokyo Prefecture)

Fig. 4 Declining *shotengai* (center of Kashiwazaki city, Niigata Prefecture)

B. Detective method of commercial accumulations

Digital yellow page data were used to detect CAs. Yellow page data have information on the business categories of each tenant. First, the business categories that construct business categories were determined. Second, selected yellow page data were converted into point data through address geocoding. Finally, polygon data of CAs were developed by an original buffering method from all point data of the yellow page data. Previous studies have also created polygons based on accumulations of buffer polygon (Sheppard et al.[4], Konishi et al.[5] and Takami et al.[6]). Details on the selection of business categories and problems of the existing buffering method are in a previous paper (Akiyama et al.[7]).

In our previous study, CAs included wholesale districts, ; however, in this study wholesale CAs were excluded to detect CAs more definite than that detected in the previous study. As a result, the selected business categories were narrowed down. In this study 401 business categories were used. The selected business categories included grocery stores, clothing stores, household and general shops, hair salons, drug stores, restaurants, sporting-goods stores, amusement facilities (e.g., video game arcades and pachinko parlors), medical clinics, real-estate offices, banks, accommodations, etc.

Polygons of CAs were developed to merge buffer polygons based on the locations of selected yellow page data by the original buffering method developed in this study. The distances of the buffer from locations of the yellow page data were defined as shown in equation 1.

$$D_N = \left(\sum_{k=1}^{n} d\min_k + d\min_N \right) \cdot (n+1)^{-1} \qquad (1)$$

where $dmin_N$ is the distance from the nearest data point to point N, n is the number of data points within R meters from point N, $dmin_n$ is the distance from the nearest data point to point n and D_N is the buffering distance of point N.

"Fig. 5" shows the calculation flow of the buffering distance. The buffering distance of point P_N in the figure is calculated. First, the nearest neighbor distances of all data points are calculated including point P_N. The nearest neighbor distance for P_N is 15.1 m. Second, the average value of the nearest neighbor distances for all data points within R meters from point P_N is calculated. This is the buffering distances of P_N (=D_N). After verifying various R distances to make appropriate polygons, 30 meters was determined as the optimal R value.

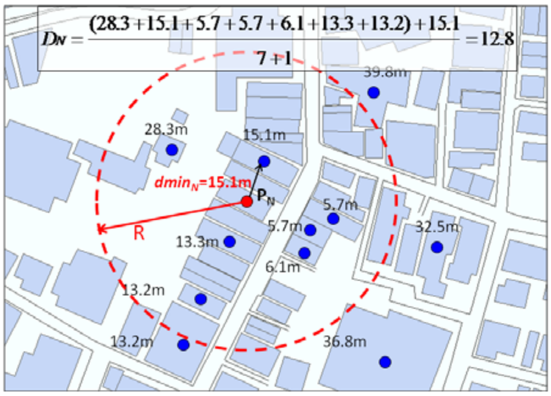


Fig. 5 Calculation of buffering distance of point P_N

Moreover data with D_N values that are too large are excluded. Therefore, buffer polygons are not created if D_N was over 100m.

In addition, following method is used if D_N is less than 100m. First, all yellow page data are allocated in 1/1000 grids of longitude and latitude based on their location information, as shown in "Fig. 6". Second, deviation scores of the buffering distance are calculated based on the buffering distances in a grid including D_N and its neighboring grids ("Fig.7 and 8"). Finally, data are excluded if the deviation scores are over 60. For the example shown in "Fig. 8", a buffer polygon was not created because the deviation score of D_N was over 60.

"Fig. 9" shows the buffering results for the proposed method. The triangle symbols in "Fig. 9" denote shops on the website for the *shotengai* located in this area ("*nodai dori shotengai*" in Japanese). Not only are almost all shops constructed in the same *shotengai* included in the same polygon, but also shops located far from the *shotengai* are not included in the CA polygons.

Finally, polygons that include a small number of data are excluded. According to commercial statistics of Japan, districts with more than 20 shops are defined as "*shotengai*". However, it cannot be expected there are more than 20 shops at clusters of roadside shops in suburban regions and small *shotengais* in local cities. Thus the threshold in this study to recognize CA was 10. In other words, polygons are not recognized as CAs if the number of data points in a polygon is less than 10.

Fig. 6 Allocation on gird

Fig. 7 Search area for calculation of deviation scores for D_N and buffering distances [m]

Fig. 8 Deviation scores based on buffering distances

C. Validation of the proposed method for cross-checking statistical data

The validity of the proposed method for developing CAs needed to be determined. To do so, locations of *shotengai* in 2004 for Tokyo Prefecture were collected into a *shotengai* directory. The directory contains the addresses of representative persons, shops, organizations, and the number of shops in a *shotengai*.

Fig. 9 Buffering result for proposed method

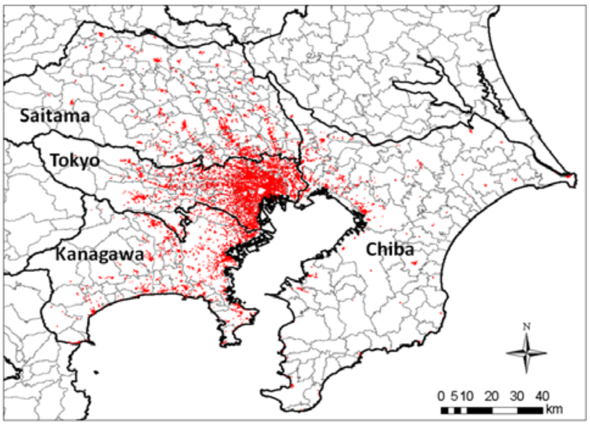


Fig.10 Distribution of CAs in South Kanto region in 2005 (red polygons)

The number of shops in this directory was used to approximate the number of shops in *shotengai*. In addition, the *shotengai'* data in the directory can be converted into point data with longitude and latitude through address geocoding, i.e., point data can be created that have the number of shops in *shotengai*.

On the other hand, the CA data are in the form of polygons. The polygon data were correlated spatially with the *shotengai'* data in the directory if latter were located in former. After integrating the *shotengai'* point data in the directory for 2004 with the CA polygons data for 2005, 1617 out of 2184 data points (74.04%) in the directory were found to be contained in the CA polygon data.

"Fig. 11" shows a correlation between the number of shops in the directory and the polygon data. When there are many points from the directory in one polygon, the number of shops in the polygon was correlated with the sum total of points contained in the polygon. The correlation coefficient was 0.6643. The polygon data clearly showed a positively correlation with the data in the directory. There results show that the proposed method is suitable for developing CA polygons.

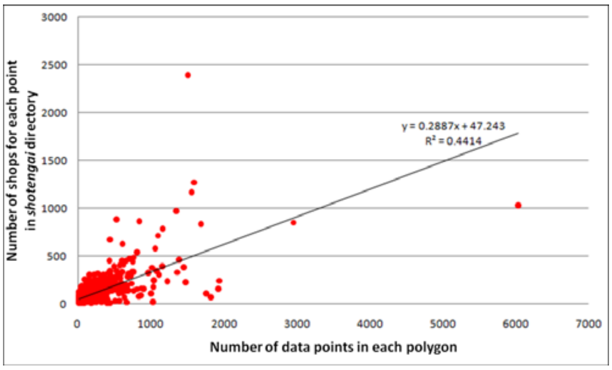


Fig. 11 Correlation between number of data points in each polygon and number of shops for each point in *shotengai* directory

IV. Analysis

A. Scale change of commercial accumulations

The number of tenants for CAs in 2000 located in the same locations in 2005 can be acquired to develop CA polygons for 2000 and 2005, and integrate them spatially. By subtracting the number of tenants in 2005 from the number of tenants in 2000, scale change, i.e., changes in the number of tenants for all CAs, can be monitored.

"Fig. 12" shows scale changes of CAs, and "Fig. 13" shows the results for each prefecture. Accumulations in almost all areas tended to reduce their scale. In contrast, the number of tenants and areas of CAs around the main train terminals in the special wards of Tokyo tended to increase. In addition, there were similar trends around main terminals outside Tokyo Prefecture. One reason that there are many accumulations to increase their scale substantially in the special wards of Tokyo in "Fig. 12"is to integrate some another accumulations in 2000 into one accumulation in 2005. However, almost all accumulations in small cities, suburbs, and rural areas stayed the same or decreased their scale.

B. Turnover rate of tenants for each commercial accumulation

Next the turnover rates of tenants in each CA were monitored. The turnover rate of tenants is defined by how how many tenants changed or emerged between the two different years. The turnover rate of tenants is defined as shown in equation 2.

$$M_i = \left(\sum C_i + \sum E_i\right) \cdot \left(\sum N_i\right)^{-1} \qquad (2)$$

where M_i is the turnover rate of tenants for a CA*i*, C_i is the number of changing tenants for a CA*i*, E_i is the number of emerging tenants of a CA*i*, and N_i is the number of tenants of a CA*i*.

The time-series information of tenants can be acquired to use the spatio-temporal data integration system introduced in section II. First, the yellow page data from 2000 is integrated spatio-temporally with yellow page data form 2005 by the

International Journal of Earth, Energy and Environmental Sciences
ISSN: 2517-942X
Vol:4, No:6, 2010

system. Second, these data points are integrated with the CA polygon data. The sum total of tenants, number of changing tenants, and number of emerging tenants can be assigned to each CA through this integration. Finally, the turnover rate of tenants for all CAs is acquired to apply equation 2 to each CA.

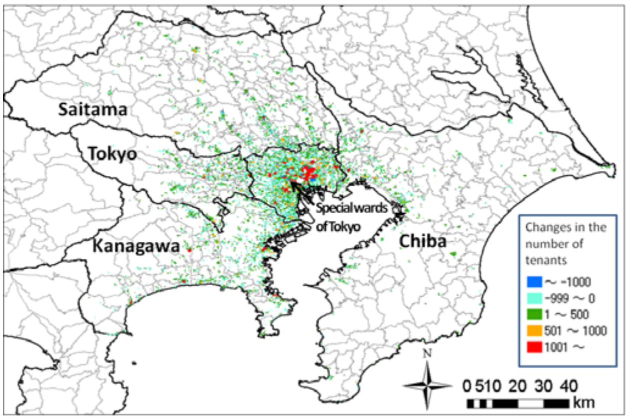
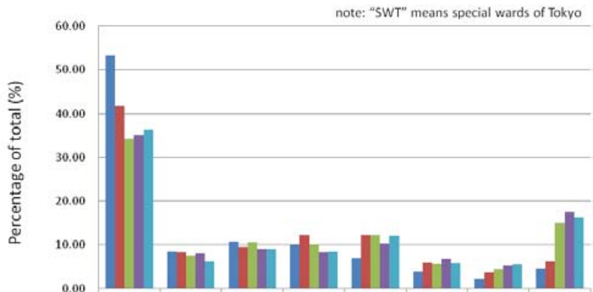


Fig. 12 Scale changes of CAs (between 2000 and 2005)



	-30 ~	-30 ~ -20	-20 ~ -10	-10 ~ 0	0 ~ 10	10 ~ 20	20 ~ 30	~ 30
SWT	53.34	8.48	10.79	9.94	6.88	3.81	2.18	4.59
without SWT in Tokyo	41.73	8.27	9.45	12.34	12.20	6.04	3.67	6.30
Kanagawa	34.16	7.58	10.60	10.11	12.33	5.67	4.50	15.04
Chiba	35.05	8.00	8.98	8.32	10.30	6.79	5.15	17.42
Saitama	36.39	6.28	9.08	8.46	12.13	5.85	5.50	16.32

Rate of scale change (%)

Fig. 13 Rate of scale changes of CAs in each prefecture

"Fig. 14" shows the turnover rate of tenant of all CAs and "Fig 15" shows same rate in the center of Tokyo. The results were interesting. Accumulations with small turnover rates of tenants are located around the central part of Tokyo in a donut shape. Accumulations with a relatively large turnover rate of tenants, though not to the extent of the central part of Tokyo, exist outside of the toroidal region around the center of Tokyo. In addition, toroidal regions similar to that around the center of Tokyo were also found in Chiba and Yokohama cities (prefectural capitals).

In this study, chain stores were extracted based on tenant names. First, name library of chain stores was developed to use the "*Telepoint Data*". The *Telepoint Data* are yellow page data published by ZENRIN CO., LTD. In this data, the business categories of major chain stores are classified as chain store names (e.g., the business category of Seven-Eleven is "Seven-Eleven", not "convenience store").

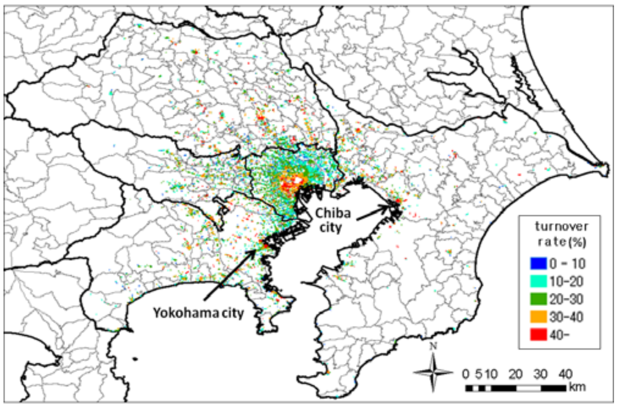


Fig. 14 Turnover rate of tenant

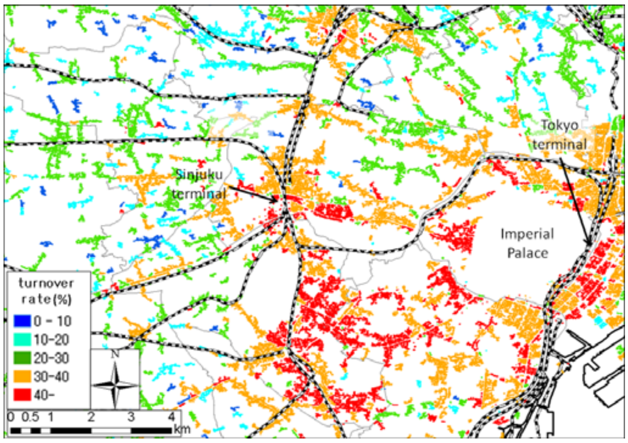


Fig. 15 Turnover rate of tenant in the center of Tokyo

The names of major chain stores and companies were collected to use this data. In addition, information from the web pages (http://www.fcajapan.gr.jp and http://jfa.jfa-fc.or.jp/) were used. Second, data for tenant names containing the names of chain store were extracted to use this library. Finally, the rate of chain stores for all CAs were calculated to count the number of chain stores in each CA and then divided by the sum totals of stores for each CA.

"Fig. 15" shows the rate of chain stores for each CA. The results clearly show that the nearer the CA is to the city center, the lower the rate of chain stores; suburban areas have the opposite trend.

The rate of chain stores was especially large for Chiba and Saitama Prefecture ("Fig. 16"). On the other hand, the rate of chain stores was small in Tokyo Prefecture. The rate for a CA with very large values, (i.e., >= 50%) was 1.34% in Tokyo Prefecture. Kanagawa Prefecture had 2.50%, Chiba Prefecture had 2.03% and Saitama Prefecture had 2.02%. Almost all CAs in this category had large shopping centers and tenants occupying these buildings. These results show that many large-scale stores open outside Tokyo Prefecture, and provide evidence of suburbanization in local cities and rural areas of the South Kanto region.

International Journal of Earth, Energy and Environmental Sciences
ISSN: 2517-942X
Vol:4, No:6, 2010

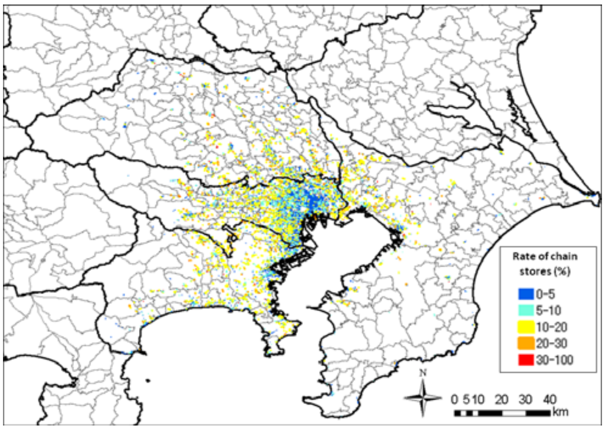
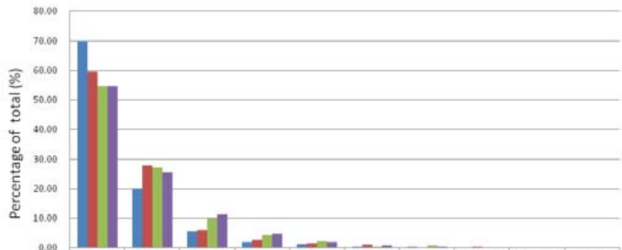


Fig. 15 Rate of chain stores for each CA



	0 ≦ n ≦ 10	10 < n ≦ 20	20 < n ≦ 30	30 < n ≦ 40	40 < n ≦ 50	50 < n ≦ 60	60 < n ≦ 70	70 < n ≦ 80	80 < n ≦ 90	90 < n ≦ 100
Tokyo	69.99	20.11	5.54	1.96	1.27	0.45	0.45	0.17	0.00	0.05
Kanagawa	59.55	27.89	6.12	2.82	1.49	1.22	0.32	0.48	0.05	0.05
Chiba	54.67	27.34	9.88	4.32	2.29	0.53	0.79	0.09	0.09	0.00
Saitama	54.58	25.52	11.25	4.83	1.95	0.94	0.50	0.29	0.14	0.00

Rate of chain stores (%)

Fig. 16 Rate of chain stores for CAs in each prefecture

D. Results

The actual states and changes of CAs were monitored with focus on the following points.

Scale change of CAs

Turnover rate of tenant in each CAs

Rate of chain stores of CAs

Scales of CAs at the center of Tokyo increased the number of tenants and area. There are many cases in the center of Tokyo where new large CAs were created by integrating multiple old CAs. Outside the center of Tokyo, some polygons increased the number of tenants without increasing their areas. This means that the tenant density increased in these areas. On the other hand, the number of tenants and area decreased in other regions.

The turnover rate of tenants in each CA showed a distinctive structure. The rate was large at the center of Tokyo and the prefectural capitals. Nearby suburban areas of the city center in the shape of a donut had small values. In addition, more suburban areas had value intermediate between city centers and the close suburban areas. The high value of turnover rate in the city centers is due to the high value of real-estate and high degree of competition for commercial activities. In contrast, the middling value of turnover rate in suburban areas far from the city center is due to increasing tenant vacancy and the opening of new chain stores.

The rates of chain stores in CAs are small in the center of Tokyo and prefectural capitals. The value increased away from the city centers. The rate is especially large in Chiba and Saitama Prefectures. This clearly shows that these prefectures are undergoing suburbanization.

However these analytical methods for CAs should be further examined. The analytical method presented in this paper is simply the first step. In future work, CAs will be analyzed with other interesting factors, and new methods will be explored to explain phenomena in CAs and integrate multiple factors.

V.CONCLUSION

The proposed method for creating CA polygons is simple and can be quantitatively modeled with high reliability. In addition, digital yellow page data also exist in foreign countries. This implies that the proposed method can be used in many other countries.

ACKNOWLEDGMENT

This work was supported by the Ministry of Land, Infrastructure, Transport and Tourism of Japan.

REFERENCES

[1] T. Ato, K. Omura, T. Arata, and S. Hujii, "The stagnation of commercial accumulation districts in front of the stations in the suburbs of the Tokyo metropolitan area: A study of Honatsugi and Odawara," *City planning review. Special issue, Papers on city planning*, vol. 41(3), 2006, pp.1037-1042.

[2] Y. Akiyama, T. Shibuki, and R. Shibasaki, "Development of three dimensional monitoring dataset for tenants variations in broad urban area by spatio-temporal integrating digital house maps and yellow page data," *IET International Conference on Intelligent Environments 08*, 2008, CD-ROM, Paper No. 0066.

[3] Y. Akiyama, and R. Shibasaki, "Development of detailed spatio-temporal urban data through the integration of digital maps and yellow page data and feasibility study as complementary data for existing statistical information," *11th International Conference on Computers in Urban Planning and Urban Management*, CD-ROM, 2009, Session8.2-2.

[4] E. Sheppard, H. Leitner, R. B. Mcmaster, and H. Tian, "GIS-based measures of environmental equity: Exploring their sensitivity and significance," *Journal of Exposure Analysis and Environmental Epidemiology*, vol9, 1999, pp.18-28.

[5] J. Konishi, T. Tamura, M. Nishime, and S. Kakumoto, "Extracting uninhabited area for regional analysis," *Papers and proceedings of the Geographic Information Systems Association*, vol.13, 2004, pp.435-438

[6] K. Takami, and Y. Sadahiro, "Change of the distribution of retail stores in retail clusters", *Papers and proceedings of the Geographic Information Systems Association* , vol9, 2000, pp.211-214.

[7] Y. Akiyama, H. Sengoku, and R. Shibasaki, "Development of time-series dataset of shops and offices through the integration of digital map and yellow page data and analysis of present state," *The 30th Asian Conference on Remote Sensing*, CD-ROM, 2009, PS3-25.