Automatic Mapping of Urban Area in High Resolution with LLGC and Integration with Existing Urban Area Maps

Hiroyuki Miyazaki*, Xiaowei Shao*, Koki Iwao**, Ryosuke Shibasaki*
* Center for Spatial Information Science, The University of Tokyo
** National Institute of Advanced Industrial Science and Technology
heromiya@csis.u-tokyo.ac.jp (Corresponding author)

INTRODUCTION

Urbanization has been a main concern for regional and global environmental change (Foley et al., 2005) and socio-economics (Angel et al., 2005). Various kinds of studies (e.g. Balk et al., 2005; Scholes and Biggs, 2005; Montgomery, 2008; Sutton et al., 2009), have used satellite-derived global urban area maps to evaluate critical aspects of urbanization for global environmental change, such as size, scale and form of cities and conversion of land cover (Laumann, 2005). The studies using global urban area map had provided valuable information about urbanization especially for less documented regions. As the studies on urbanization progressed, however, 1-km spatial resolution of global urban area map have gotten obsolete to measure spatial structure of urban area in fine scale (Angel et al., 2005) and to model land use conversion with socio-economical variables (Nelson and Robertson, 2007).

To measure spatial structure of urban area in fine scale, urban area map have to be developed for each case from high-resolution satellite images (e.g. Landsat, Terra/ASTER, IKONOS and Quickbird). However, classification of urban area from high-resolution satellite images is much time-consuming process, preventing not only effective progress of studies on urbanization but also international comparison. Thus we believe that developing and providing high-resolution global urban area map would contribute to deeper understanding of urbanization.

In this paper, we present automatic algorithm for developing global urban area map from high-resolution satellite images using Learning with Local and Global Consistency (LLGC) technique and integration with existing urban area maps using logistic regression. Regarding definition of urban, we introduced definition by Potere and Schneider (2007), who defined urban with presence of built-up area.

METHOD

We constructed the method with two steps: first, we classified urban area from high-resolution satellite images using Learning with Local and Global consistency (LLGC) technique; second, to correct misclassification of LLGC, we integrated the urban area map of LLGC with existing urban area map of coarse resolution. The procedure in each step is described below; overview of the method is showed in Figure 1.

Classifying Urban Area from ASTER/VNIR Satellite Images using LLGC

To achieve automatic classification of urban area from satellite images, the two basic components of classification, clustering and labeling, have to be automated. For automated clustering, unsupervised clustering method (e.g. ISODATA) is commonly employed for land cover classification (e.g. Koeln et al., 2000; Angel et al., 2005); for labeling, however, visually interpreting clusters into land cover classes is needed because the clusters does not have any information of land cover class.

For automated labeling, we employed urban area map of coarse resolution as training data. Successfully classified urban area map would be a good training data for clustered satellite images;

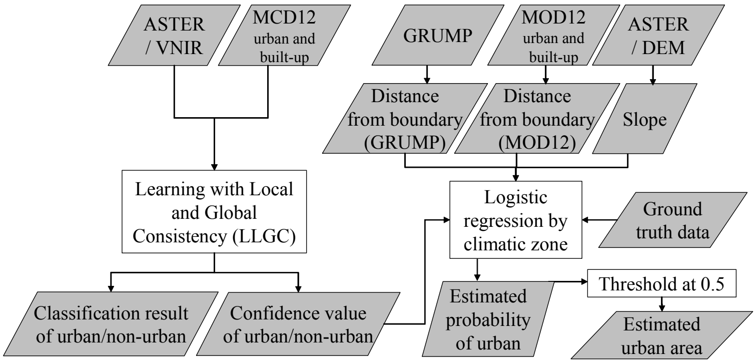

Figure 1: Flowchart of processing for high-resolution urban area map.

however the gap of spatial resolution makes inconsistency among pixel values and labels. For example, if a cluster which is likely to be urban and another cluster which is likely to be non-urban are covered within a urban pixel of coarse resolution, each cluster includes training data of urban even though they should be separated into urban and non-urban.

To deal with the gap of spatial resolution, we introduced Learning with Local and Global Consistency (LLGC). LLGC constructs a function to correct roughly labeled classification into smoothly labeled result (Zhou et al., 2003). The method was thus suitable for our case, in which clusters derived from high-resolution satellite images were initially labeled with coarse-resolution urban area map.

The output of LLGC has not only classification result, but also confidence value ranged from 0 to 1. The confidence value was rather appropriate to represent ambiguous gradation between urban and rural area. In order to deal with the gradation, we introduced the confidence value into integration mentioned below.

We employed surface reflectance images derived from Visible and Near-Infrared Radiometer of Advanced Spaceborne Thermal Emission and Reflection radiometer (ASTER/VNIR), which have been commonly used for monitoring urban environment. 15-m spatial resolution of ASTER/VNIR is much finer than existing global urban area map, thus urban area map derived from ASTER/VNIR would allow measuring complex spatial structure of urban area. ASTER/VNIR has been operated since December in 1999 to complete cloud-free global coverage (Yamaguchi et al., 1998); therefore we supposed that ASTER/VNIR was the most suitable source for high-resolution global urban area map.

Integrating with Existing Urban Area Map using Logistic Regression

Although the clusters were successfully classified owing to LLGC, the result would include misclassifications due to similarity in surface reflectance among different land covers. For example, urban area surrounded with sand area could be classified as non-urban because surface reflectance of urban is similar to sand area. Cloud cover would considerablly lead to misclassification. These disturbances stem from heterogeneities of landscape and quality among ASTER/VNIR scenes, suggesting the result would have uncertainty by scene.

To reduce the uncertainties among scenes, we integrated the result of LLGC with existing urban area maps by introducing logistic regression. Logistic regression has an advantage in representing presence of urban in form of probability, which could repsent spatial gradation between urban and rural area.

We also considered geographical heterogeneity in accuracy existing urban area maps. Schneider et al. (2003) had suggested that, in satellite-based estimation of urban area, the accuracy at urban centre is higher than at urban fringe. We expected that distance from boundary of urban area (DBU) would work as proxy of the heterogeneity. We calculated DBU from urban class cluster of MODIS/Terra Land Cover Type 96-Day L3 Global 1km ISIN Grid V004[1] (MOD12) and GRUMP Urban Extent Grid[2] (GRUMP).

Terrain is also significant factor for presence of urban (Clarke et al., 1997). Therefore we included slope calculated from digital terrain model (DEM) derived from ASTER/VNIR into the logistic regression.

We constructed a model to estimate probability of presence of urban as equation (1).

$$P_i(urban) = \frac{\exp(U_i)}{1+\exp(U_i)} \qquad (1)$$

where $P_i(urban)$ is the probability of presence of urban at ith pixel and U_i is defined in form of polynomial expression as equation (2).

$$U_i = \beta_0 + \beta_1 \times LLGC_i + \beta_2 \times SLOPE_i + \beta_3 \times DBU_{MOD12,i} + \beta_4 \times DBU_{GRUMP,i} \qquad (2)$$

where β is coefficient for each variable, $LLGC_i$ is confidence value of LLGC at ith pixel, $SLOPE_i$ is slope at ith pixel, $DBU_{MOD12,i}$ and $DBU_{GRUMP,i}$ is DBU at ith pixel in MOD12, GRUMP, respectively (positive value for outside urban area; negative value for inside urban area).

EXPERIMENT AND RESULT

Sampling ASTER/VNIR Scenes and Ground Truth Data

For experiment, we sampled ASTER/VNIR images by following group:

Group A: Randomly selected scenes under stratification in terms of number of cities by climatic zone (7 scenes for tropical; 13 for dry; 55 for temperate; 24 for cold).
Group B: Scenes intersecting with urban area of city of more than one million (241 scenes).

On each scene of Group A, approximately 500 point coordinates were sampled at lattice grid; whereas, on the scenes of Group B, 799 point coordinates were sampled from GRUMP/Settlement Point database and 83 point coordinates were sampled from Degree Confluence Project database (Iwao et al., 2006). We acquired ground truth data by visually interpreting presence of urban at each point coordinates using false color composites of ASTER/VNIR image based on color tone and texture.

Classifying Urban Area from ASTER/VNIR using LLGC

We excluded pixels assigned as water body and vegetation cover, at which Normalized Difference Water Index (NDWI) were higher than 0 or Normalized Difference Vegetation Index (NDVI) were higher than 0.5, and then processed clustering analysis on the ASTER/VNIR images

[1] http://duckwater.bu.edu/lc/mod12q1.html
[2] http://sedac.ciesin.columbia.edu/gpw/

using ISODATA method in which initial number of clusters was 100 and tolerant convergence was 2%.

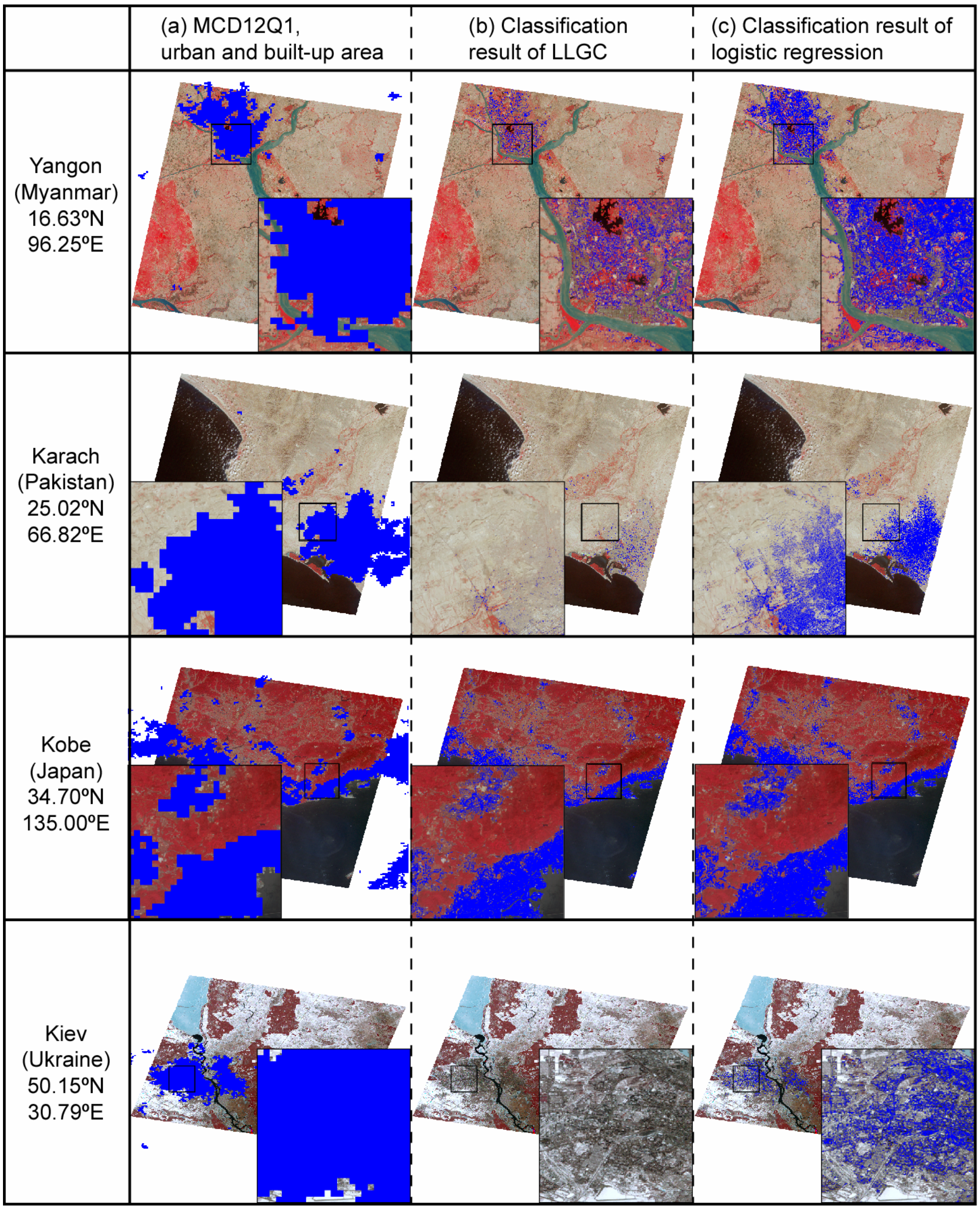

Figure 2: Examples of estimated urban area of MCD12Q1, result of LLGC, and result of logistic regression. Square in image represent extent of close-up image.

We applied LLGC method to the clustered images using MODIS Terra + Aqua Land Cover Type Yearly L3 Global 500 m SIN Grid[3] (MCD12Q1) as initial label of urban area. The result represented spatial structure of urban area, such as sparse greenness in urban area and gradient in building

---

[3] https://lpdaac.usgs.gov/lpdaac/products/modis_products_table/land_cover/yearly_l3_global_500_m/mcd12q1

between urban and rural area, in much finer resolution than MCD12Q1 (Column (a) and (b) in Figure 2). However, we found that no urban area was classified in the scene even though urban area was visually recognizable (Column (b) for Karach and Kiev in Figure 2).

Accuracy assessment on the result of LLGC classification showed 79% user's accuracy, 47% producer's accuracy, 91% overall accuracy, and 0.62 kappa coefficient (Table 2Table).

Estimating Probability of Presence of Urban using Logistic Regression

We assigned confidence value of LLGC, slope, DBU of MOD12 and GRUMP (negative value for

Variable	Tropical	Dry	Temperate	Cold
Intercept	-2.237 (< 0.001)	-2.548 (< 0.001)	-2.024 (< 0.001)	-2.254 (< 0.001)
Real-valued classification with LLGC	6.737 (< 0.001)	6.772 (< 0.001)	4.724 (< 0.001)	5.460 (< 0.001)
Slope	0.010 (0.301)	-0.016 (0.361)	-0.046 (< 0.001)	-0.049 (< 0.001)
Distance from boundary of MOD12	-25.287 (< 0.001)	-13.795 (< 0.001)	-18.997 (< 0.001)	-5.018 (< 0.001)
Distance from boundary of GRUMP	-6.215 (0.003)	-8.854 (< 0.001)	-5.696 (< 0.001)	-2.693 (< 0.001)

Table 1: Coefficients and p-values of logistic regression by climatic zone. Number in each row indicates coefficient for the variable; number in brackets indicates p-values.

	Climatic zone	User's accuracy	Producer's accuracy	Overall accuracy	Kappa coefficient
LLGC	Global	79%	40%	94%	0.50
	Tropical	81%	36%	93%	0.46
	Dry	81%	22%	96%	0.33
	Temperate	79%	43%	93%	0.53
	Cold	79%	38%	94%	0.49
MOD12	Global	59%	61%	93%	0.56
	Tropical	65%	54%	93%	0.55
	Dry	45%	60%	95%	0.49
	Temperate	58%	65%	92%	0.57
	Cold	61%	56%	93%	0.55
GRUMP	Global	29%	85%	80%	0.35
	Tropical	37%	91%	84%	0.46
	Dry	25%	85%	88%	0.34
	Temperate	30%	87%	78%	0.34
	Cold	27%	80%	80%	0.32
Estimated urban area with logistic regression	Global	78%	50%	94%	0.58
	Tropical	75%	60%	94%	0.64
	Dry	81%	43%	97%	0.54
	Temperate	78%	52%	94%	0.59
	Cold	77%	44%	94%	0.53

Table 2: Result of accuracy assessments on LLGC, MOD12, GRUMP and estimated urban area with logistic regression globally and by climatic zone.

inside urban area; positive value for outside urban area) to ground truth data, and estimated logistic model for probability of presence of urban by climatic zone[4] (Table 1). Signs of coefficients were corresponded to our assumption and statistically significant at 99% level, except coefficient of slope for tropical zone and dry zone.

We calculated probability of presence of urban for each pixel and classified the pixels into urban or non-urban. Pixels of more than 0.5 probability were classified as urban ; pixels of less than or equal to 0.5 probability were classified as non-urban (Column (c) in Figure 2). We assessed accuracy of the result of the classification, showing 78% user's accuracy, 57% producer's accuracy, 91% overall accuracy, and 0.92 kappa coefficient; whereas the accuracies of MOD12 were 71% user's accuracy, 60% producer's accuracy, 91% overall accuracy and 0.90 kappa coefficient; the accuracies of GRUMP were 29% user's accuracy, 85% producer's accuracy, 80% overall accuracy, 0.35 kappa coefficient (Table 2). Higher overall accuracies and kappa coefficients of the results of logistic regression than those of the others indicate improvement in accuracy owing to the integration.

DISCUSSION

Evaluation of LLGC Classification

Higher user's accuracy of the result of LLGC than producer's accuracy of that indicates that LLGC failed to detect much part of actual urban area (Table 2). The main causes of the failures would be so similar surface reflectance among urban and the other non-vegetated land cover that much urban area was classified as non-urban.

Despite of the misclassifications of LLGC, complex spatial structure of urban area, which was filled with a few pixels in coarse-resolution urban area maps, was very finely represented (Figure 2Table). Similar extent to MCD12 urban area map indicates that LLGC worked as trimmer on MCD12 urban area map, suggesting that coarse-resolution urban area map could be improved in terms of spatial resolution with keeping original geographical distribution.

Implication on Coefficients of the Logistic Regression

Difference in degree of the coefficients among climatic zones reflects features of source data. For example, coefficients of DBU of MOD12 for tropical and temperate zone were more significant than that for dry and cold zone. It indicates that, for tropical and temperate zone, estimated urban area in MOD12 was more closely associated to actual urban area than for dry and cold zone. The explanation could be supported with the accuracy assessment, in which kappa coefficients of MOD12 for tropical and temperate zone were higher than that for dry and cold zone.

Improvement with the Integration

The urban area map classified with the integration was more accurate than the result of LLGC, MOD12 and GRUMP in terms of overall accuracy and kappa coefficient (Table 1). Thus we might conclude that the integration improved the accuracy of urban area maps as a whole. Moreover, the integrated map might inherited the features of output of LLGC and MOD12; that is, user's accuracies of the integrated map (77-81%) were close to that of LLGC (79-81%) whereas producer's accuracies of the integrated map (43-60%) were close to that of MOD12 (54-61%).

---

[4] Climatic zone for a scene was assigned by overlaying World Map of the Köppen-Geiger climate classification (available at http://koeppen-geiger.vu-wien.ac.at/).

We visually found significant improvements in the result of Karach and Kiev, in which LLGC could not detect some extent of actual urban area, but the integration had done (Column (b) and (c) for Karach and Kiev). The accuracy assessment reflects the improvement, showing higher overall accuracy and kappa coefficient of result of the integration than that of LLGC. If LLGC had failed to detect urban area, the integration would strongly depended on MOD12 and GRUMP; however effect of confidence value of LLGC were remained in the result, shaping complex spatial structure of urban area in 15-m resolution.

CONCLUSION

We presented the method for automatic development of urban area map in high resolution using LLGC technique and by integration with existing urban area maps using logistic regression. We implemented the method with automatic algorithm, and demonstrated the method on 340 scenes of ASTER/VNIR as high resolution image, MCD12, MOD12 and GRUMP as existing urban area maps. The result showed LLGC worked effectively to trim up 500m-resolution clusters of urban area into 15m-resolution clusters. The result also showed the integration using logistic regression improved accuracy of urban area maps better than LLGC-derived urban area map and existing urban area maps.

The proposed method will be practically useful for improving accuracy and spatial resolution of global urban area maps. The high-resolution global urban area map developed with the method will encourage providing deeper insights on urbanization not only for developed countries but also for developing countries through regionally and internationally comparative studies.

ACKNOWLEDGEMENT

This research used ASTER Data beta processed by the AIST GEO Grid from ASTER Data owned by the Ministry of Economy, Trade and Industry of Japan.

BIBLIOGRAPHY

Angel, S., S. C. Sheppard, et al., 2005 The Dynamics of Global Urban Expansion. Transport and Urban Development Department, The World Bank, pp 200.

Balk, D., A. Storeygard, et al., 2005 Child hunger in the developing world: An analysis of environmental and social correlates. Food Policy, 30(5), 584-611.

Clarke, K., S. Hoppen, et al., 1997 A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. Environ. Plann. Plann. Des., 24, 247-262.

Foley, J. A., R. DeFries, et al., 2005 Global Consequences of Land Use. Science, 309(5734), 570-574.

Iwao, K., K. Nishida, et al., 2006 Validating land cover maps with Degree Confluence Project information. Geophys. Res. Lett., 33, L23404.

Koeln, G. T., T. B. Jones, et al., 2000 GeoCover LC: Generating Global Land Cover from 7600 Frames of Landsat TM Data. Proceedings of ASPRS 2000 Annual Conference.

Laumann, G., 2005 Science Plan: Urbanization and Global Environmental Change. Bonn, Germany, International Human Dimensions Programme on Global Environmental Change, pp 61.

Montgomery, M. R., 2008 The Urban Transformation of the Developing World. Science, 319(5864), 761-764.

Nelson, G. C. and R. D. Robertson, 2007 Comparing the GLC2000 and GeoCover LC land cover datasets for use in economic modelling of land use. Int. J. Remote Sens., 28(19), 4243-4262.

Potere, D. and A. Schneider, 2007 A critical look at representations of urban areas in global maps. GeoJournal, 69(1), 55-80.

Schneider, A., M. A. Friedl, et al., 2003 Mapping Urban Areas by Fusing Multiple Sources of Coarse Resolution Remotely Sensed Data. Photogramm. Eng. Rem. S., 69(12), 1377–1386.

Scholes, R. J. and R. Biggs, 2005 A biodiversity intactness index. Nature, 434(7029), 45-49.

Sutton, P. C., S. J. Anderson, et al., 2009 Paving the planet: impervious surface as proxy measure of the human ecological footprint. Prog. Phys. Geog., 33(4), 510-527.

Yamaguchi, Y., A. B. Kahle, et al., 1998 Overview of Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER). IEEE Trans. Geosci. Remote Sens., 36(4), 1062-1071.

Zhou, D., O. Bousquet, et al., 2003 Learning with Local and Global Consistency. Adv. Neural Inf. Process. Syst., 16, 321-328.