# Scale Verification in GyberGIS: A Case Study in Road Change Detection

Jin Xing, Renee Sieber

McGill University, Montreal, Quebec, Canada
Email: jin.xing@mail.mcgill.ca; renee.sieber@mcgill.ca

Keywords: CyberGIS, Scale Verification, Road Change Detection, Geospatial CyberInfrastructure

## Introduction

Big data has complicated land use/cover change (LUCC) research but also opens opportunities for Geospatial CyberInfrastructure (GCI) (Wang et al., 2013), which is an effective approach to handle big data. The sheer volume of datasets shifted the analysis from single computer to cloud computing (Yang et al., 2011) because data size now far exceeds the capacity of individual computers. Variety in data formats and spectral bands can demand different sets of analysis tools to support each type (e.g., to account for band-sharpened imagery). The velocity of big data enables multi-temporal LUCC, while increasing problems due to resolution heterogeneity (both spatial and spectral) and creating difficulties in ground truthing. Because GCIs propose to enable knowledge discovery, we believe that big data enabled-GCI will attract considerable interest in LUCC.

Scale is illustrative of the challenges in using big data for LUCC. For us, geographic scale is defined as remotely sensed image resolution and the geospatial extents of imagery. With the ever-increasing spatial and spectral resolutions afforded by RS platforms, it is very difficult to guarantee all the RS images are scale-homogeneous. Resolution heterogeneity will shift LUCC research from pixel-based methods to the object-based approaches (Chen et al., 2012). Additionally, geospatial extent will play a pivotal role in image decomposition and recomposition. To optimize distributed computing of big LUCC datasets, the geospatial extent will likely vary in each data chunk according to prevalence of objects under study (Xing, Sieber, and Kalacska, 2014). Scale not only impacts LUCC analysis methods, but also the underlying computation process, especially in the era of big data.

## Scale in Road Change Detection and LUCC-GCI

The accuracy of road extraction and change detection depends on the scale of the RS imagery datasets. For example, it will be very difficult to extract roads from Landsat 8 images even with spatial interpolation techniques. Landsat 8 imagery is offered at 30m spatial resolution and most roads are narrower than 30m. Figure 1 illustrates road extraction, based on the graph cut and Support Vector Machines (SVM) classification techniques (Song and Civco, 2004). In Figure 1(c), the road features nearly disappear because the high image resolution mixes the road with other types of features (e.g., cars and trees). From Figure 1(d), the coarser resolution helps extract major roads but cars and trees still obscure some roads, especially at the right bottom corner. In Figure 1(e), numerous false changes are caused by the cars, trees, and differences in road pavement. Manually determining suitable scales for road extraction and change detection is onerous and error-prone, so we look to integrate the scale verification into LUCC-GCI for this task. The meaning of scale verification is two-fold: checking if the scales of imagery datasets are appropriate and selecting the right scale factors for image scaling and change identification.

We propose a hybrid parallel-serial workflow to better integrate scale in a LUCC-GCI. Our LUCC-GCI is composed of the LUCC layer, the resource allocation layer, the dataflow management layer, the computation layer, and a domain layer. The LUCC layer contains the code for pre-processing, image change identification, and accuracy assessment. The resource allocation layer controls the scheduling of different tasks in our LUCC-GCI. The dataflow management layer implements decomposition and recomposition method as the big data analysis workflow. At last, the computation layer integrates cloud computing and Hadoop framework to provide scalable resource provisioning, and data storage management (Almeer, 2012). Some of the processing will occur in parallel (e.g., image classification and change identification) and some will be serial. A domain layer inserts into the GCI algorithms and other treatments to identify the meaning of changes (e.g., when is a road not the lane between parked cars in a parking lot?). At the beginning of each recursion, the domain layer is provided feedback (e.g., execution time, resource utilization, change detection accuracy, and the corresponding image scales) from the other layers, and updates the GCI with fine-tuned domain-specific parameters and algorithms. The initial domain knowledge is quite critical for the accuracy of LUCC and the number of recursions for to handle the multitude of possible scales. Including data with inappropriate scales and selecting the wrong scale baseline will create lags in the processing of the big heterogeneous LUCC data.

**Results and Discussion**

We investigate road change detection in the Greater Montreal Area, Canada, from 2005 to 2015 to demonstrate the importance of scale verification in GCI. The data used is listed in Table 1. We build a test bed for our case using the private cloud at McGill University and virtual machines from Microsoft Azure cloud. Initial domain-specific parameters are created from sampling and testing, which will be updated in the recursions mentioned above. These include removing vehicles and people since, depending on the resolution they are not important for road identification. We add data elimination rules, to remove datasets providing very little information or too much noise due to the scale heterogeneity. Image scaling methods are implemented to address the challenge of scale-heterogeneity for image change identification. The image-scaling algorithm we choose is a combination of discrete wavelet transformation and bicubic interpolation, and the baseline is selected as 0.11m spatial resolution (0.11m is the finest scale in our data). Within each recursion, the domain layer checks the utility of the datasets, updates the image change identification parameters, and adjusts the image scaling factors.

Table 2 shows the accuracy of road change detection from the six recursions of scale verification. The accuracy is calculated by using 200 ground truth points. We notice the Landsat 8 data was eliminated at the first recursion because very few roads can be extracted from it. Although we can manually eliminate Landsat 8 data to reduce the computation time, here we want to prove the effectiveness of the scale verification in LUCC-GCI. The accuracy generally increases as we downscale the spatial resolutions. The low accuracy at the beginning of the recursions are caused by the spatial interpolation of DMTI datasets. Choosing another scale baseline coarser than 0.11m will likely avoid these low accuracy outputs, at the risk of missing the useful scales. The research indicates the highest accuracy coming with 1.76m for road change detection. Integrating scale verification workflow into a GCI can help us address the scale heterogeneity challenge afforded by big-data enabled LUCC.

**Acknowledgements**

This research is funded by Microsoft Azure Research Award.

**References:**

Almeer, M. H. (2012). Cloud hadoop map reduce for remote sensing image analysis. *Journal of Emerging Trends in Computing and Information Sciences*, 3(4), 637-644.

Celik, T. (2009). Multiscale change detection in multitemporal satellite images. *Geoscience and Remote Sensing Letters*, IEEE, 6(4), 820-824.

Chen, G., Hay, G. J., Carvalho, L. M., & Wulder, M. A. (2012). Object-based change detection. *International Journal of Remote Sensing*, 33(14), 4434-4457.

Green, K., Kempka, D., & Lackey, L. (1994). Using remote sensing to detect and monitor land-cover and land-use change. *Photogrammetric engineering and remote sensing*, 60(3), 331-337.

Miller, H. J., & Goodchild, M. F. (2015). Data-driven geography. *GeoJournal*,80(4), 449-461.

Radke, R. J., Andra, S., Al-Kofahi, O., & Roysam, B. (2005). Image change detection algorithms: a systematic survey. *Image Processing*, IEEE Transactions on, 14(3), 294-307.

Song, M., & Civco, D. (2004). Road extraction using SVM and image segmentation. *Photogrammetric Engineering & Remote Sensing*, 70(12), 1365-1371.

Wang, S., Anselin, L., Bhaduri, B., Crosby, C., Goodchild, M. F., Liu, Y., & Nyerges, T. L. (2013). CyberGIS software: a synthetic review and integration roadmap. *International Journal of Geographical Information Science*, 27(11), 2122-2145.

Yang, C., Goodchild, M., Huang, Q., Nebert, D., Raskin, R., Xu, Y., ... & Fay, D. (2011). Spatial cloud computing: how can the geospatial sciences use and help shape cloud computing?. *International Journal of Digital Earth*,4(4), 305-329.

Xing, J., & Sieber, R. E. (2016). A land use/land cover change geospatial cyberinfrastructure to integrate big data and temporal topology. *International Journal of Geographical Information Science*, 30(3), 573-593.

Xing, J., Sieber, R., & Kalacska, M. (2014). *The challenges of image segmentation in big remotely sensed imagery data*. *Annals of GIS*, 20(4), 233-244.

**Table 1. Details about dataset used in the road change detection**

| Year | Platform | Spatial Resolution (m) | Spectral Resolution |
|---|---|---|---|
| 2005 | Montreal Metropolitan Community Orthophotos | 0.11 | RGB (sharpened & fused) |
| 2006 | DMTI | 0.6 | RGB (sharpened & fused) |
| 2007 | Montreal Metropolitan Community Orthophotos | 0.3 | RGB (sharpened & fused) |
| 2009 | DMTI | 0.6 | RGB (sharpened & fused) |
| 2012 | DMTI | 0.6 | RGB (sharpened & fused) |
| 2015 | Landsat 8 | 30 | 11 bands (0.43mm-12.51mm) |

Datasets are provided by DMTI Spatial, Communauté métropolitaine de Montréal, and USGS EROS, respectively.

**Table 2. Road change detection accuracy under different spatial resolutions.**

| Scale (m) | TT (%) | TF (%) | FT (%) | FF (%) | Overall Accuracy (%) |
|---|---|---|---|---|---|
| 0.11 | 18.0 | 40.5 | 27.0 | 14.5 | 32.5 |
| 0.22 | 18.0 | 31.0 | 35.0 | 16.0 | 34.0 |
| 0.44 | 27.0 | 28.0 | 31.0 | 14.0 | 41.0 |
| 0.88 | 33.0 | 16.5 | 10.0 | 40.5 | 73.5 |
| 1.76 | 45.0 | 5.5 | 8.5 | 41.0 | 86.0 |
| 3.52 | 40.5 | 8.5 | 8.0 | 43.0 | 83.5 |

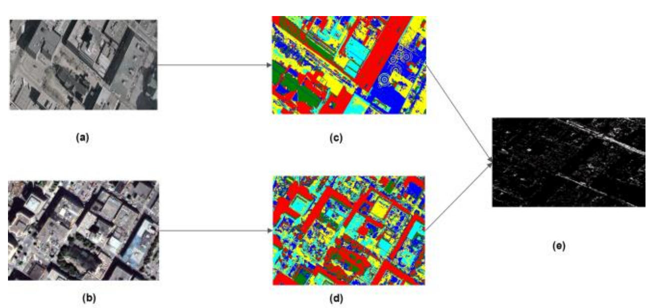


Figure 1 (a) Sample from Montreal Metropolitan Community Orthophotos with 0.11m spatial resolution at 2005; (b) sample from DMTI image with 0.6m spatial resolution at 2006; (c) is the classification result from (a); (d) is the classification result from (b); (e) is the change map generated from (c) and (d), which contains lot of false changes.