Multi-level Change Detection in Urban Areas Using ZY-3 Multi-temporal Stereo Imagery

Dawei Wen[1], Hui Liu[1]

1. The State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing
Wuhan University
Wuhan, China
Email: daweiwen@whu.edu.cn

Jiayi Li[2], Xin Huang[2,1]

2. The School of Remote Sensing and Information Engineering
Wuhan University
Wuhan, China

Abstract—**Due to the rapid process of urbanization, there is an urgent need for accurate and detailed change detection in urban areas. Ziyuan-3 (ZY-3), as China's first civilian high spatial resolution satellite, carrying the three-line array camera, enables the acquisition of in-track stereo images. The characteristics of multi-angle imaging and high-spatial resolution are fully exploited to explore subtle urban change information. In this study, a multi-level (pixel-grid-block) change detection framework using stereo images was conducted to detect information related to land cover transitions, changed hotspots, and multi-temporal landscape. The experimental results reveal some interesting and informative findings: (1) our proposed multi-level method is effective in detecting change details, with kappa coefficient reaching 0.81 at the pixel level and correctness being 95% at the grid level; (2) the block-level landscape analysis indicate that there was greater fragmentation and spatial heterogeneity of building landscape during the urbanization process; (3) the mean nearest neighbor distance between building patches decreases by about 1.0 m during 2012 and 2013.**

I. INTRODUCTION

URBANAZATION is taking place at an unprecedented rate around the world in terms of scale and pace. The remote sensing techniques hold the great potential for monitoring urban changes. A large set of works have been dedicated to urban monitoring by utilizing coarse or moderate resolution data, e.g. MODIS and Landsat [1-3]. However, the subtle urban dynamics, such as building construction or removal, which are more important for urban planning and management, can be neglected in the previous work. Fortunately, the increase of spatial resolution in satellite imagery enables us to detect subtle changes at a very fine scale.

Despite the advantages of the improved spatial resolution, change detection using high-spatial resolution images suffers from some problems, e.g., uncertainty of the spectral information (i.e., intra-class variance increase and inter-class variance decrease) [4] and spatial heterogeneity in the multi-temporal domain (e.g., spatial mis-registration, parallax distortion for high architectures, different viewing angles) [5].

Consequently, higher resolutions do not guarantee superior detection capability. In this regard, there have been techniques proposed for improving change detection capability. On one hand, it is widely acknowledged that the incorporation of spatial features can improve the interpretation of high-spatial resolution images [6, 7]. On the other hand, the stereo capability of recent satellite sensors, such as IKONOS, WorldView, and ZY-3, make it possible to generate multi-temporal orthographic images, so as to minimize the spatial inconsistency as much as possible. Considering the above mentioned aspects, we attempted to incorporate the characteristics of multi-angle imaging and high-spatial resolution to explore subtle urban change information.

In this study, we proposed a multi-level (pixel-grid-block) change detection framework using multi-view ZY-3 satellite data. At the pixel level, the detailed change trajectories are detected. Secondly, the pixel-based land cover maps are aggregated to the grid level, which can highlight the changed hotspots and reduce the pepper and salt effect. Finally, the landscape composition and configuration at the block level are analyzed to indicate urban landscape spatiotemporal change patterns.

II. STUDY AREA AND DATASETS

The study area is the downtown area of Wuhan, Hubei. It covers an area of approximately 375 km^2. Wuhan, situated at the conjunction of Yangtze River and Han River, is one of the biggest metropolises in the central China, and has favorable geographical advantages of linking the eastern and western regions. With these geographical advantages, Wuhan has been well known as "thoroughfare to nine provinces", which provides favorable conditions for rapid development of the city. In this study, the ZY-3 images of Wuhan acquired in 2012 and 2013 were used. The orthographic image for the study area is shown in Fig. 1.

III. METHODS

The objective of our proposed framework is to analyze change information related to land cover trajectories, hotspots,
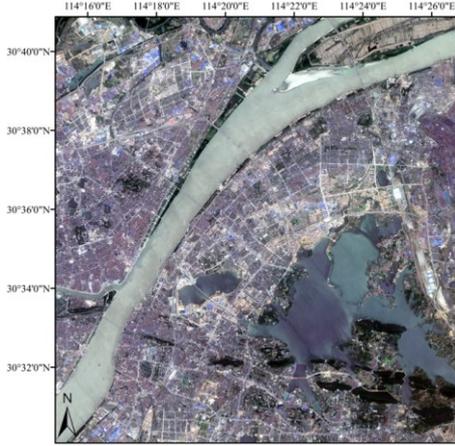
Fig. 1. Orthographic image for the study area of Wuhan.

and landscape during the urbanization process. The key steps of the proposed method are: (1) orthographic image and DSM generation; (2) building height retrieval; (3) multi-temporal land cover mapping; (4) multi-level change detection.

A. Orthographic image and DSM generation

The fundamental for automatic generation of orthographic image and DSM using multi-view imagery is image matching. Here we adopted semi-global matching [8] as dense matching technique in the photogrammetric derivation. Semi-global matching was considered due to good tradeoff between runtime and performance.

B. Building height retrieval

DSMs are potential to provide the height values of the main urban structures, i.e., buildings. Here, the performance of building height retrieval from ZY-3 stereo imagery was evaluated in order to determine how to use the DSMs in the following change detection procedure. Firstly, the normalized DSMs (nDSMs), describing height attribute of land surface features, is calculated using top-hat filter by reconstruction [9]. Subsequently, the random forest classifier that combines spectral bands, normalized difference vegetation index, morphological building/shadow index [10, 11] and nDSM is applied for building extraction. Finally, estimated building heights from nDSM are validated by actual height obtained from the urban planning department.

C. Multi-temporal land cover mapping

To guarantee the effectiveness and efficiency of the proposed change detection framework, the multi-temporal land cover mapping framework that integrates multi-classifier ensemble approach and multi-temporal sample mitigation strategy is proposed. For land cover mapping framework, each classifier with stacked spectral bands and different spatial features (i.e., textural, morphological, and multiple indexes features) is fused at the decision level, the details of which can be refereed in [4]. In order to further refine the initial result, semantic relationship between different land cover categories (e.g., spatial adjacency between buildings and shadow) and larger height attributes for
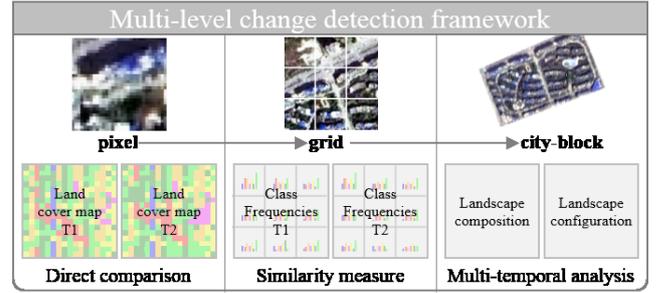

Fig. 2. Framework for multi-level change detection.

above-ground structures (i.e., buildings and trees) are considered in the post-processing steps. Additionally, it should be noted that there is a great deal of time and labor devoted to samples collection in each classification in multiple time series. To resolve this problem, a sample migration strategy, reusing sampling positions that were not changed during the time span, is proposed. In this way, only one set of training samples is required for the whole multi-temporal land cover mapping.

D. Multi-level change detection

The multi-level (pixel-grid-block) change detection framework (Fig. 2) aims at analyzing change information in different aspects. At the pixel level, land cover trajectories are analyzed by focusing "from-to" information for each pixel. At the gird level, changed areas (i.e., changed hotspots) are focused by analyzing the class frequencies in a local region. At the city-bock level, landscape metrics (both landscape composition and landscape configuration) are calculated to reveal general trend of landscape evolution in each city-block.

At the pixel level, land cover trajectories can be acquired through post-classification comparison based on multi-temporal land cover maps. Post-classification comparison method can produce "from-to" change information by directly comparing class labels pixel by pixel. At the grid level, each grid is further divided into a series of non-overlapping cells (3×3 cells in this paper); and the frequencies of land cover types in each cell are used to characterize land cover component and its spatial distribution [12]. In this way, the significantly changed areas (i.e., hotspots) can be distinguished in terms of land cover and their spatial arrangement. It should be noted that the grid size of 51 pixels was considered suitable and reasonable due to the best balance between change detection accuracy and details preservation. At the city-block level, a set of metrics, i.e., building coverage ratio (BCR), largest patch index (LPI), edge density (ED), mean and standard deviation of patch area (MPA and SPA), mean shape index (MSI), mean and standard deviation of nearest neighbor distance (MNN and SMN), patch density (PD), and cohesion index (CI) are adopted to measure landscape composition and configuration [13]. Here, the city-block networks are generated from road networks (both primary and secondary) of open street map. It should be noted that each metric is calculated as city-block being the basic unit, thus making the local landscape pattern in urban areas being fully understood.

	LPI	ED	MPA	SDPA	MSI	SSI	MNN	SMN	PD	CI
WH12	34.1	644	0.28	0.99	1.69	1.28	10.8	7.7	342	97.7
WH13	32.5	699	0.21	0.79	1.59	1.17	9.8	6.6	448	97.3

IV. RESULTS AND DISCUSSION

A. Accuracy for building height retrieval

For accuracy assessment of the building height retrieval, 400 building were randomly selected. By comparing the building heights estimated by nDSM and their actual values, some interesting conclusions can be summarized: (1) nDSM serves

as an important feature in the supervised buildings extraction procedure with feature importance being 0.37; (2) ZY-3 derived nDSM can achieve satisfactory performance for building height smaller than 50 m with root mean square error below 4 m; (3) estimated values for building height larger than 50 m can be seriously underestimated due to large disparity and occlusion for high buildings in the epipolar image.

B. Pixel-level change detection result

For accuracy assessment of pixel-level change detection result, we computed a confusion matrix (Table I) using 30 randomly selected testing points per change trajectory. Generally, satisfactory performance can be achieved with overall accuracy being 84.4% and Kappa coefficient reaching up 0.81. Additionally, producer's accuracy and user's accuracy for most change trajectories are larger than 85.0%. The satisfactory accuracy indicate that our prosed multi-temporal land cover mapping framework with sample migration strategy is effective.

C. Grid-level change detection result

The result of grid-level change detection is shown in Fig. 3. The map shows that changes occur both in urban core and fringe, suggesting rapid urbanization accommodating with infrastructure construction and urban renovation. Furthermore, the changed hotspots are assessed quantitatively. Firstly, the map of changed hotspots was acquired by setting a threshold via minimum cross entropy [14]. Subsequently, identified changed hotspots are checked via careful visual inspection. For quantitative accuracy assessment, 1885 blocks out of 1984 ones are correctly detected, achieving a satisfactory correctness of 95.0%.

D. Block-level change detection result

For landscape composition analysis, maps of building coverage ratio (BCR) are shown in Fig. 4, as well as some representative examples for significantly changed city-blocks. As can be seen from the figure, most city-blocks with high BCR values are located in the old city area along the Yangtze River, indicating dense buildings covered in these city-blocks.
The changes of BCR values are related to demolition of old urban villages (Fig. 4(a)) or residential buildings (Fig. 4(b)), and new buildings construction (Fig. 4(c)).

For landscape configuration analysis, the mean values for a series of configuration metrics are calculated (Table II). A
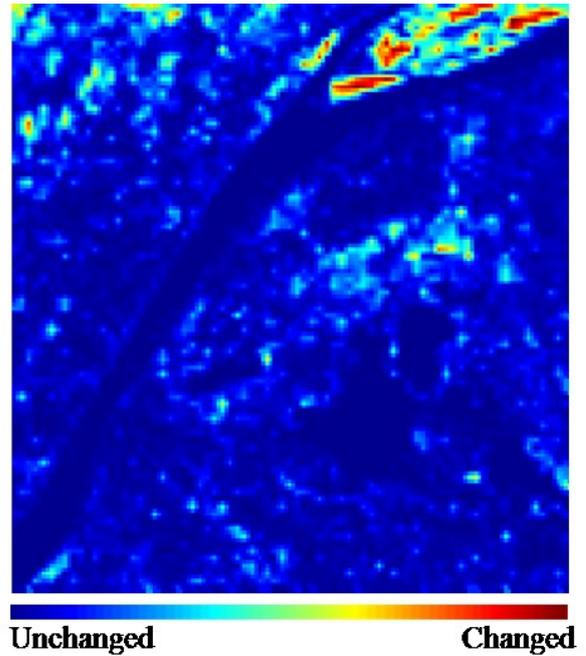

Unchanged   Changed

Fig. 3. Result of grid-level hotspot detection.

decrease of the LPI coupled with increase of the PD is indicative of greater fragmentation and more spatial heterogeneity. Meanwhile, MSI retained a slight downward trend; the value decreased approximately 0.1, revealing smaller shape complexity for the newly constructed buildings. The patch area related metrics, i.e., MPA and SDPA, are decreased. This finding can be attributed to smaller-than-average buildings (e.g., detached and high apartments) construction or bigger-than-average buildings (e.g., dense and compact residential area) demolition during urbanization process. Moreover, the measure of average nearest neighbor distance, MNN, decreased 1.0 m from 2012 to 2013.

V. CONCLUSIONS

In this study, a novel multi-level (pixel-grid-block) change detection framework using ZY-3 stereo imagery was proposed. The pixel, grid, and city-block level analysis focus on change trajectories, hotspots pattern, and landscape evolution. Thus, a comprehensive and systematic monitoring is ensured to guide the urban management and planning. At the pixel level, the integration of multi-classifier ensemble, samples migration, and post-classification comparison is presented to derivate change trajectories. At the grid level, hotspots are identified to analyze the spatial pattern of urban renovation and the quantitative accuracy assessment was conducted. At the block level, both landscape composition and configuration are analyzed. The proposed framework was successfully implemented in the downtown area of Wuhan with satisfactory performance.

TABLE I
CONFUSION MATRIX FOR PIXEL-LEVEL CHANGE DETECTION RESULT

| Classified data	Reference data						Total	User's Acc. (%)
	soil->buil	buil->soil	soil->grass	grass->soil	water->grass	nochange		
soil->buil	23	0	0	0	0	0	23	100.0
buil->soil	1	23	0	0	0	0	24	95.8
soil->grass	0	0	26	0	0	0	26	100.0
grass->soil	0	0	0	20	0	0	20	100.0
water->grass	0	0	2	0	30	0	32	93.8
nochange	6	7	2	10	0	30	55	54.6
Total	30	30	30	30	30	30	180	
Producer's Acc. (%)	76.7	76.7	86.7	66.7	100.0	100.0		
Overall classification accuracy (%)		84.4						
Kappa coefficient		0.81						
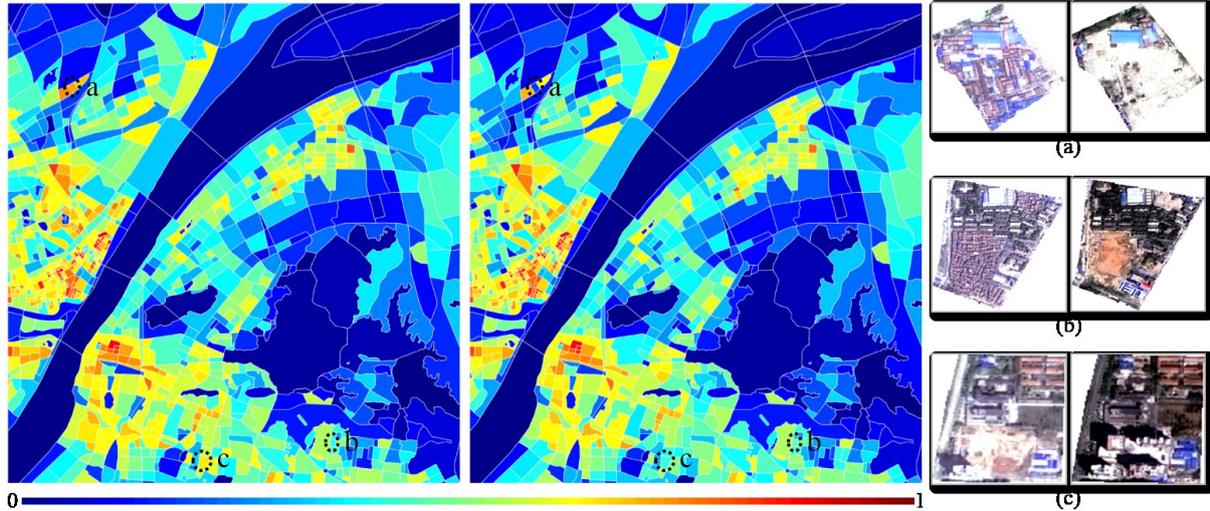
Fig. 4. Multi-temporal analysis of building coverage ratio with some representative cases ((a)-(c)).

REFERENCES

[1] L. Wang, S. Wang, W. Li, and J. Li, "Monitoring urban expansion of the Greater Toronto area from 1985 to 2013 using Landsat images," in *2014 IEEE Geoscience and Remote Sensing Symposium*, 2014, pp. 4264-4267.

[2] P. Gamba, G. Lisini, G. Trianni, and E. Angiuli, "Urban area mapping using ASAR wide swath mode and Landsat data: A comparison on eastern China," in *Urban Remote Sensing Event (JURSE), 2013 Joint*, 2013, pp. 127-130.

[3] C. M. Mertes, A. Schneider, D. Sulla-Menashe, A. Tatem, and B. Tan, "Detecting change in urban areas at continental scales with MODIS data," *Remote Sensing of Environment,* vol. 158, pp. 331-347, 2015.

[4] X. Huang and L. Zhang, "An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery," *Geoscience and Remote Sensing, IEEE Transactions on,* vol. 51, pp. 257-272, 2013.

[5] A. Stumpf, J.-P. Malet, P. Allemand, and P. Ulrich, "Surface reconstruction and landslide displacement measurements with Pléiades satellite images," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 95, pp. 1-12, 2014.

[6] X. Huang, Q. Lu, and L. Zhang, "A multi-index learning approach for classification of high-resolution remotely sensed images over urban areas," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 90, pp. 36-48, 2014.

[7] R. Dianat and S. Kasaei, "Change detection in optical remote sensing images using difference-based methods and spatial information," *IEEE Geoscience and Remote Sensing Letters,* vol. 7, pp. 215-219, 2010.

[8] H. Hirschmuller, "Accurate and efficient stereo processing by semi-global matching and mutual information," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, pp. 807-814.

[9] R. Qin and W. Fang, "A hierarchical building detection method for very high resolution remotely sensed images combined with DSM using graph cut optimization," *Photogrammetric Engineering & Remote Sensing,* vol. 80, pp. 873-883, 2014.

[10] X. Huang and L. Zhang, "A Multidirectional and Multiscale Morphological Index for Automatic Building Extraction from Multispectral GeoEye-1 Imagery," *Photogrammetric Engineering & Remote Sensing,* vol. 77, pp. 721-732, 2011.

[11] X. Huang and L. Zhang, "Morphological building/shadow index for building extraction from high-resolution imagery over urban areas," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of,* vol. 5, pp. 161-172, 2012.

[12] D. Wen, X. Huang, L. Zhang, and J. A. Benediktsson, "A Novel Automatic Change Detection Method for Urban High-Resolution Remotely Sensed Imagery Based on Multiindex Scene Representation," *Geoscience and Remote Sensing, IEEE Transactions on,chaoli* vol. 54, pp. 609-625, 2016.

[13] K. McGarigal and B. J. Marks, "FRAGSTATS: spatial pattern analysis program for quantifying landscape structure," 1995.

[14] C. H. Li and C. Lee, "Minimum cross entropy thresholding," *Pattern Recognition,* vol. 26, pp. 617-625, 1993.