CHANGE DETECTION BASED ON SPATIAL DATA MINING

LI Deren [a], XIA Song [b,] *, SUI Haigang [a], ZHANG Xiaodong [a]

[a] National Laboratory for Information Engineering in Surveying, Mapping & Remote Sensing, Wuhan University, China

[b] Information School, Zhongnan University of Economics and Law, Wuhan, China – xiasongwh@yahoo.com.cn

KEY WORDS: Photogrammetry, Land Cover, Change Detection, Data Mining, Photography

ABSTRACT:

Spatial data mining (SDM) and change detection (CD) are hot issues in the field of geospatial information processing. Considering the feasibility of discovering knowledge from spatial database and the demands of knowledge for change detection, the paper presents that change detection needs knowledge and technology of SDM should be integrated into applications of change detection. Special knowledge for change detection of linear feature, area feature and terrain are studied. Relationship between accuracy of change detection and errors of image registration is discussed. Change detection of linear feature, area feature and terrain based on SDM are investigated respectively. Change detection based on multivariate statistical analysis and training samples are analyzed. Land use/cover change detection based on SDM is discussed.

1. INTRODUCTION

With technical progress in technology of spatial data acquisition, mass storage and network interconnection, volume of spatial data has been increasing dramatically. Vast data satisfied potential demands of exploring the earth's resource and environment by human being, widening exploitable information source, but the processing approaches of spatial data lag behind severely, and are unable to discover relation and rules in large amount of data efficiently and make full use of existing data to predict development trend. Due to lack of advanced means to find knowledge hidden behind, it has become an imminent bottleneck that data are excessive while knowledge is scarce. For geospatial data, the ability of production and transmission is far greater than the ability of interpretation and analysis (Li Deren etc, 2006). Demands on novel effective methods for extracting attractive knowledge and information from various spatial database or warehouse are more and more urgent. SDM was proposed under the weight of application requirements. And it provides significant ways for discovering valuable knowledge to realize automation and intelligence of processing spatial data.

On the other hand, with construction and variation of all kinds of spatial database, a growing attention on timely data updating and change detection has been paid to. Self-driven real-time on-board change detection techniques are needed. Therefore, key problems such as automatic image registration, image matching, feature extraction, object interpretation, image fusion and data cleansing need solving. And automatic spatial data mining and knowledge discovering (SDM&KD) also need to be resolved in order to construct intelligent change detection system (Li Deren, 2002).

2. CHANGE DETECTION NEEDS SDM&KD

Prior knowledge could be provided for selecting source data and approaches of change detection by research on reasons and discipline of terrain change. It is of instructive significance for investigation on effective change detection method to analyze origins of change, study rules of change, and explore relationship of image variation and terrain change.

There are three categories of change reasons: abiological action in nature, biological action, and human activity among which human activity is most notable factor. Especially, research on change detection of urban area is focused on recently. Many theories and methods on change detection of urban area is put forward, lots of indices and models for ranking urban development are devised, and they contribute to proper planning and advance alert of cities greatly. Terrain changes are synthetic reflect of three kinds of changes spatially and temporally. Complexity of terrain change severely affects the automation of change detection.

Change forms on multi-temporal remotely sensed imagery include positional deviation, radiant variation and spatial structure alteration. It comprises actual relief change and other imaging factors. Study on change detection is to decrease and weaken difference caused by other non-object change and acquire real change information.

From above analysis, it can be concluded that (1) Terrain change is regular, regularity needs to be studied, and knowledge database may come into being, which can help to change detection; (2) Changes especially resulted from human activity can help to change detection using planning information; (3) As for changes caused by imaging factors, 3D, multi-source, pixel level and feature level integrated detection approach; (4) SDM can discover further knowledge to aid change detection.

3. KNOWLEDGE FOR CHANGE DETECTION

Actually, spatial data are inclined to high autocorrelation. Knowledge types discovered from spatial database include general geometric knowledge, object oriented knowledge, spatial characteristic/discriminate rules, spatial classification/regression rules, spatial clustering/association rules, spatial dependent/predictable rules, spatial serial rules, spatial exceptions and outliers (Li Deren etc, 2006). All sorts of knowledge interrelates each other, and multiple rules are required for given occasions.

3.1 General Change Detection Knowledge

Generally speaking, knowledge can be regarded as information structure formed by association of related information (Wang Yongqing, 1998). As for change detection, there are two kinds of knowledge for exploitation. (a) Context knowledge, i.e. about the context and objects therein when the image is acquired. (b) Course knowledge, i.e. about what knowledge, when to use, how to use, etc.

Context knowledge includes:
(1) Object spectrum characteristic.
(2) Feature and evolution rules of spatial or temporal distribution.
(3) Knowledge on models of sensors.
(4) Knowledge on objective geographical environment.
(5) Knowledge on imagery.
(6) Object property, spatial relation and association rules.
(7) Expert's interpretation knowledge.
(8) Geosciences auxiliary data.

In order to make full use of limited computer resource, it is very important to ascertain what to do and how to do in each processing step. So it is necessary to determine the next strategy using course knowledge which is also known as control knowledge. As for change detection, course knowledge involves multi-level object extraction strategy, multi-scale detection strategy, multi-source data integration strategy, and other new theory or methods in the field of cognitive psychology etc. Several controlling strategies can be used in change detection as follows (Sui Haigang, 2002):

(1) Grouping is a crucial content in mid-level computer vision, and it is also a strategy in common use.

(2) Easiest and most reliable objects take priority of others in extraction.

(3) According to image resolution, proper extraction algorithm should be chosen.

(4) Multi-level extraction approach. Extraction should be performed using radiant or geometric information at first, then topological network relation, and advanced top-level knowledge at last.

(5) Multi-scale detection approach. Large-scales can exhibit global feature or objects, and extraction and confirmation of detailed objects can be realized on small-scales.

(6) Fusion and collaboration of multi-source data, multiple clues and varied information. Introduction of new information or data can decrease difficulty of detection or extraction, but merging or fusing becomes another complicated problem.

3.2 Special Knowledge for Change Detection of Line Feature or Area Feature

In order to identify and detect geospatial linear feature or area feature exactly, it is inevitable to describe and comprehend practical linear objects accurately and model them properly. As to characteristics of linear feature, many scholars have published their research findings (Gary P, Steve W., 2000), and they can be concluded as follows:

(1) Geometric property
For linear objects, it mainly includes width, width consistency, figure, size and orientation, curvature, link mode etc. For area objects, it mainly includes area, perimeter, compactness, divergence, ductibility, rectangle similarity, orientation, curvature etc。

(2) Radiometric property
It means reflecting trait and contrast comparing with surroundings of linear objects or area objects.

(3) Topologic property
It expresses completeness degree of linear objects as a part of the network. Mostly, topologic properties of roadway and water system are focused on. A common descriptor for area objects is Euler number E=N-H, where N represents number of connected parts and H represents number of holes.

(4) Context property
It mainly includes correlative property of around area and local context clues.

(5) Functional property
Functions born by linear objects or area objects in objective world and therefore other relative properties can be inferred.

3.3 Special Knowledge for Terrain Change Detection

Applications of terrain data are generally classified as two sorts. One is direct application, i.e. regarding DEM as vital component of digital mapping and elements of geospatial database. The other is user-oriented indirect application, i.e. transforming DEM to various derived products that satisfy all kinds of specialized requirements (Li Zhilin, Zhu Qing, 2003). Terrain knowledge that can be used in change detection is summarized as follows:

(1) Primary terrain factors
Slope/aspect.
Area/volume.
Slope change rate/aspect change rate.

(2) Topographic feature
Topographic feature means significant surface points, lines or areas that can represent terrain details properly, and they make up the relief backbones. Terrain feature points include peaks, valleys, neks etc. Terrain feature lines include ridges, valley lines etc. Terrain feature areas include concavity or convexity, and are often relative to curvature of perpendicular orientation.

(3) Hydrologic analysis
It mainly comprises analysis of catchments, stream network, and drainages.

(4) Complicated terrain factors
Terrain complexity index (TCI) (Wang Lei etc, 2004), terrain granulation degree, ravine density, landform curvature, terrain fractal index (TFI), hypsography degree etc.

4. INTEGRATED CHANGE DETECTION BASED ON SDM

4.1 Relationship between Accuracy of Change Detection and Errors of Image Registration

Dai systematically investigated and quantitatively evaluated the effects of misregistration on the accuracy of remotely sensed change detection. It is interesting to notice that, among the seven Landsat TM bands, band 4 (near-infrared channel) is the most sensitive to misregistration when change detection is concerned. The results from false change analysis indicate a substantial degradation in the accuracy of remotely sensed change detection due to misregistration. It is shown that a registration accuracy of less than one-fifth of a pixel is required to achieve a change detection error of less than 10% (Dai Xiaolong, 1998).

On the basis of theory of spatial statistics, quantitative description and analysis about effect of errors of image registration on accuracy of change detection are put forward. It shows that the 5th band of Landsat, 3rd band of SPOT4 and 4th band of Quickbird are most severely affected on the same accuracy of change detection. At five different experimental regions, for identical image type and equal accuracy of change detection, there are much information in urban area and suburban area where result of change detection is more sensitive to errors of image registration. Applying the 5th band of Landsat7 and panchromatic image of SPOT4 in experiment, statistical data show that registration error of 1 pixel can make the change detection accuracy decrease to about 45%. In order to achieve change detection accuracy of 90%, image registration error of less than 0.22 pixels is required (Zhang Xiaodong, 2005). Traditionally, image registration is an important premise of change detection, and image registration with high precision is required before change detection. Aiming to the question, synchronized registration and detection algorithm is put forward.

4.2 Change Detection of Linear Feature Based on SDM

Road detection and extraction with integration of GIS and RS is depicted as: Candidate road segments are found and selected according to vector map. As for image with low resolution, buffer area can be generated with road center lines and certain distance. As for image with high resolution, buffer area can be generated with road center lines and parallel lines of road sides. All parallel lines inside the buffer area are searched, tracked and connected.

Automatic road grouping according to whole relation can be broken up into two parts. One is automatic grouping of similar road segments. It means grouping road segments one by one based on various similarity measurements in a given (usually small) area, and then the whole connection is performed. The other is extended road grouping. It means making up for the information that was omitted when we processed parallel lines.
Key parameters describing roads are extracted and formalized by fuzzy theory. Knowledge database can be established and include various object knowledge and GIS knowledge which can be represented by fuzzy production rules and predicate logic. And object identification and extraction can be realized by knowledge reasoning. Object models and parameters can be quantized by fuzzy theory, knowledge can be expressed by confidence measure, and object's quantized description can be inferred by uncertainty algorithm of confidence measure. Besides the road's properties can be the evidence, much knowledge in the knowledge database can be the proofs. As roads have wonderful network property, road intersections derived from existing roads are important proofs, and rivers and vehicles that are verified are also reliable proofs.

Experiment of the procedure is illustrated as Figure 1 and Figure 2. Figure 1 show original aerial imagery overlaid with vector roadway, which covers some country region in Hubei Province. And white bold lines are original vector graphics. Figure 2 shows results of automatic detection and extraction algorithm. Red circles mark obvious errors or omissions, black bold lines are unchanged roads, and others are new roads.

Figure 1. Original image overlaid with vector roadway

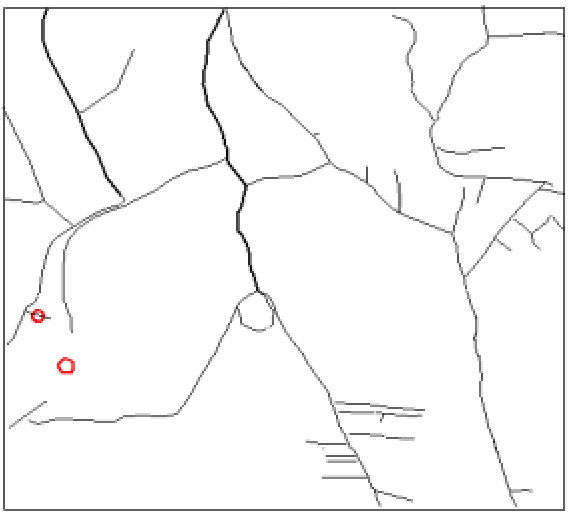

Figure 2. Detected and extracted road network

4.3 Change Detection of Area Feature Based on SDM

After rasterization of vectors are done, corresponding pixels and GIS objects are related. On a certain scale, edge extraction of remotely sensed imagery is performed, and registration of vector and imagery is realized based on extracted line feature and geometric registration model. And registration is an iterative process during which mismatching of corresponding objects is deleted and expected registration with higher precision in the next iteration. The result comprises of well registered objects which are also unchanged ground objects. Therefore geometric model parameters with high precision are obtained. As for deleted GIS objects, their polygons or center lines are used as primary models to extract area objects in corresponding local image region. And extracted objects are compared with corresponding objects in GIS data to determine whether changed or not. After several iterations, all interested GIS objects are processed and changed areas are output.

Experiment is illustrated as Figure 3, 4 and 5. Figure 3 shows Quickbird image with image resolution 0.6m and pre-orthorectification from Shanghai Pudong Zone in 2002. The region consists of farmlands, ponds, vegetation, rivers, houses etc, and is comparatively flat. Figure 4 shows pre-processed corresponding GIS data in 1998 with local georeference system. Figure 5 shows change detection results of total solution procedure, and the numbered area objects are changed objects after the period of time.

Figure 3. Quickbird image from study area in 2002

Figure 4. Pre-processed corresponding vector data in 1998

5. TERRAIN CHANGE DETECTION BASED ON SDM

Elevations of ground objects are always presumed to be constant for previous theory or applications of change detection in remote sensing. But these presumptions have not been examined, and if elevation of terrain surface has changed assuredly, unexpected trouble could effect on the result of 2D change detection (Xia Song, 2006). Li Deren believes that existing 4D products need to be fully exploited in change detection in order to augment predictive knowledge to improve efficiency and automation; image registration and change detection should be performed synchronously in order to reduce impact of registration error on results; changes of terrain elevation need to be considered for change detection of large scale spatial database (Li Deren 2002).

Figure 5. Change detection result of area objects

With existing database of control points or DEM/DOM, by image matching and spatial resection, orientation elements of new stereoscopic pair enhanced can be calculated. VLL matching is employed to interpolate elevation of grid points of entire stereo model, and the elevations are compared with old DEM. If accuracy of some point is eligible, it is an unchanged point. If difference of elevation exceeds acceptable ranges, it is a dubious point. During the process, multiple terrain factors or other related knowledge can be used to help to decide a certain ground point has changed or not. After editing and quality checking, changed points can be discriminated from dubious points, and unchanged points can take part in next iteration. Ultimate products include results of change detection, updated DEM and orientation elements of new stereo pair. If new imagery is acquired by digital camera with POS system, orientation elements provided by POS may be regarded as initial value, and it is advantageous for faster computation.

In the process of producing digital orthoimage by digital photogrammetry workstation, two orthoimage generated respectively from left and right photograph of the same stereo model should be coincident in theory, i.e. there should be no parallax. If parallaxes exist, orientation elements are precise assuredly, and there is no problem in the course of orthoimage matching, then the parallaxes occurring on orthoimage pair reflect errors of DEM which is used to produce the orthoimage pair. According to Prof. Kraus' research, elevation measuring accuracy of stereo orthoimage pair is trebly higher than elevation accuracy of DEM used for producing the stereo orthoimage pair. Therefore, traditional process of updating DEM by digital photogrammetry workstation, during which three orientation steps and lots of manual work are inevitable, can be avoided. Parallax occurring on orthoimage pair can be employed to refine outdated DEM directly to realize change detection and updating.

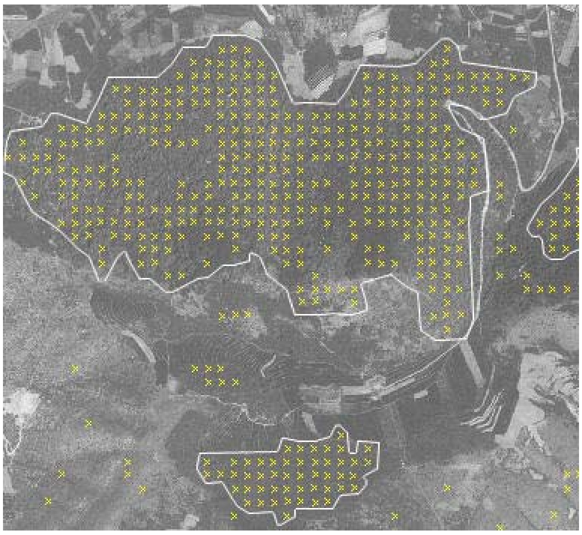

Figure 6. Terrain changed area of Hanzhong region

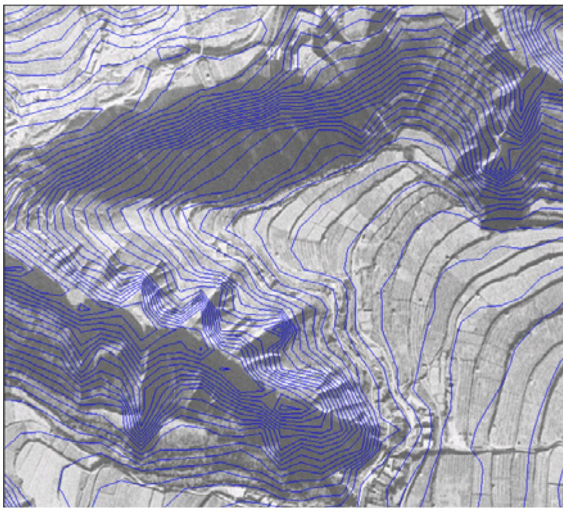

Figure 7. Overlay of contour from updated DEM and Orthoimage (Contour interval: 5m)

In order to have further investigations on the approach of terrain change detection, source materials in 1999 and 1988 of Hanzhong Prefecture in Shanxi China, including aerial imagery, DEMs, DOMs, are used for experiments. Terrain data (DEM) is produced with digital photogrammetric workstation by experienced operators in Shanxi Bureau of Surveying & Mapping, and is regarded as standard data to evaluate the presented approaches. Figure 6 shows the result of terrain changed area of Hanzhong region with change detection based on orthoimage matching. Obvious changed areas have been detected and marked on the image. Figure 7 shows overlay of contour from updated DEM and Orthoimage with contour interval 5m. It can be seen that contours are well overlaid on the surface of terrain. So the updated DEM is of high accuracy and able to satisfy practical applications.

6. CHANGE DETECTION BASED ON MULTIVARIATE STATISTICAL ANALYSIS

To eliminate impact of inter-channel correlations, canonical correlation analysis (CCA) and the so-called multivariate alteration detection (MAD) method based on CCA are introduced into bi-temporal multi-channel change detection. According to MAD method, two multi-channel imageries covering the same geographic location and acquired at different times are taken as two sets of random variables, then MAD transformation is performed on these random variable sets to produce a set of result variates that are uncorrelated with each other. In this way, correlations between channels can theoretically be removed as much as possible, so that the actual changes in all channels can be simultaneously detected in the resultant difference image (Zhang Lu, 2004).

To improve effectiveness of the MAD result, it is proposed to use signal-to-noise ratio (SNR) instead of variance as a measurement for change information distribution, and another multivariate statistical transformation called minimum noise fraction (MNF) is introduced as a post-processing step for MAD transformation. In this way, change information can be separated from noise to the greatest extent, so that the technical problem of effectively concentrating change information and producing difference image could be solved properly.

Figure 8. 2000-11-1 (Landsat7 ETM+)

Figure 9. 2002-11-8 (SPOT5 HRG)

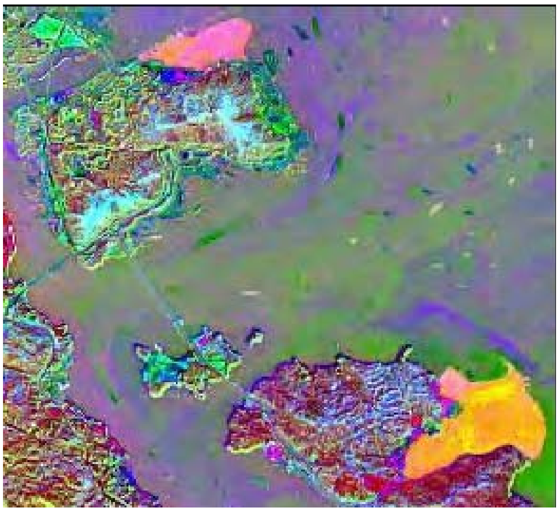

Figure 10. False color composite imagery from the 1st, 2nd, and 3rd variates of MNF/MAD

Figure 8 and Figure 9 show original imagery of test area in west Hong Kong. ETM+ is false color composite imagery of TM5, 4, and 3. HRG is false color composite imagery of XS4, 3, and 2. Figure 10 shows the 1st, 2nd, and 3rd variates of MNF/MAD.

Experimental results in a few test sites indicate that MNF/MAD method based on CCA is able to extract change information effectively from multi-temporal multi-channel remotely sensed imageries and pool them into a few resultant components of the temporal difference image. Generally these components could manifest some clear physical meanings. A distinguished advantage of the MNF/MAD scheme is its invariant to linear scaling, which means it is insensitive to disagreement in measurement scale, gain settings in measuring devices, and linear radiometric distortions; as a result the requirement on image preprocessing could be reduced.

7. CHANGE DETECTION BASED ON SAMPLES LEARNING

A novel kernel change detection algorithm (KCD) is proposed. The input vectors from two images of different times are mapped into a potential much higher dimensional feature space via a nonlinear mapping, which will usually increase the linear margin of change and no-change regions. Then a simple linear distance measure between two high dimensional feature vectors is defined in features space, which corresponds to the complicated nonlinear distance measure in input space. Furthermore the distance measure's dot product is expressed in the combination of kernel functions and large numbers of dot product processed in input space by combined kernel tactic, which avoids the computational load. Finally this paper takes the soft margin single-class support vector machine (SVM) to select the optimal hyper-plane with maximum margin.

Supervisory learning methods are advisable, considering the facts that change samples and no-change samples are asymmetry badly and change samples are acquired difficulty and no-change samples are infinite, while the soft margin single-class support vector machine (SVM) is taken in our research, translating the choice of threshold into the selection of the optimal hyper-plane with maximum margin. Choosing the suitable kernel function and parameter, we get excellent detection accuracy.

Figure 11. Image of Iraq presidential palace before assault

Figure 12. Image of Iraq presidential palace after assault

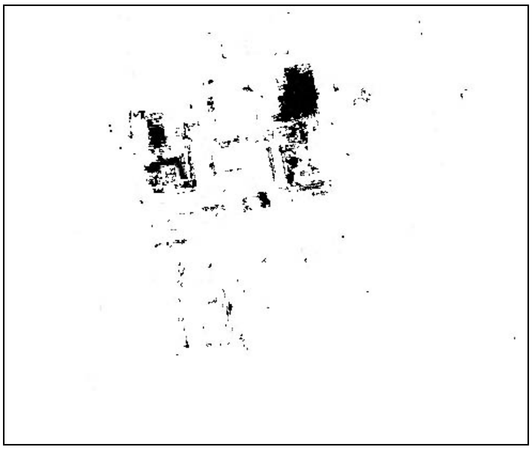

Figure 13. Result of change detection

Figure 11 and Figure 12 respectively show image of Iraq presidential palace before and after assault. Figure 13 shows change detection result based on samples learning.

8. LAND USE/COVER CHANGE DETECTION BASED ON SDM

ANN has strong ability of non-linear mapping, good self-adaptability and low demand for data distribution. Image grey levels, texture, relief information and other geodata can be integrated to realize research on land use/cover change detection based on artificial neural network (ANN) as Figure 14 to 17 show (Xiao Ping, 2001). From vision effect, land type change is firstly represented as grey level change, and texture can only accessorily discriminate partial types such as town and plantation in agriculture lands. So change detection should use imagery's grey levels, texture, or other information selectively and orderly in order not to impair precision of change detection. Proper procedure describes as follows. Imagery should be pre-processed necessarily at first such as geometric rectification, radiometric correction and spectral similarity analysis. Then primary result of change detection is achieved after ANN based change detection, and improvement can be made according to result of elementary detection and spectral similarity analysis. At last, wavelet texture featured quantity and other auxiliary geodata can help to validate change of land types.

Figure 14 and Figure 15 show two temporal imagery of the same area. Figure 16 and Figure 17 show change detection result and changed area.

Figure 14. Imagery taken in 1992

Figure 15. Imagery taken in 1996

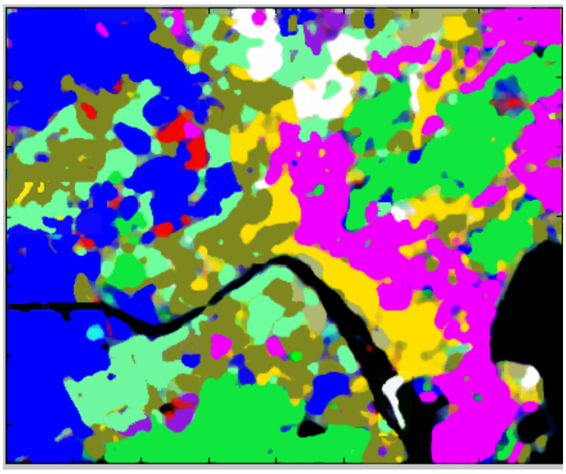

Figure 16. Change detection results

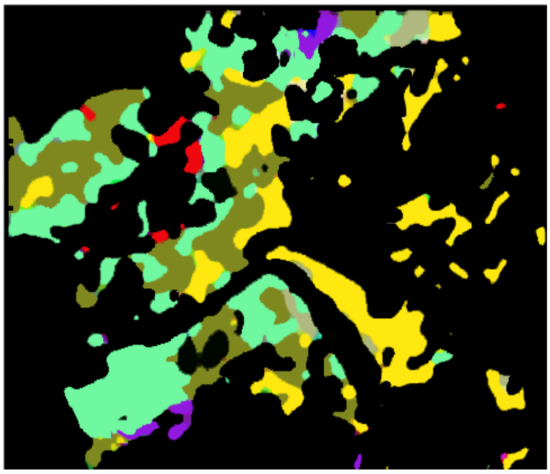

Figure 17. Changed area

9. CONCLUSIONS

Change detection is a complicated process influenced by multiple factors. Knowledge plays an important role in the process. So change detection based on spatial data mining open up prospects for research on land use/cover change. Spatial data mining & knowledge discovering (SDMKD) can explore much effective knowledge in geospatial database to assist the process and validate the result of change detection. Change detection of linear feature, area feature and terrain based on SDM are separately investigated, and elementary experiments are illustrated in the paper. Various semantic and non-semantic information works on different steps. Diverse geodata contributes to integrating different types of data to automatic and intelligent change detection.

Later-on research includes further effective use and integration of special context knowledge and course knowledge in change detection of various spatial features.

REFERENCES

Dai Xiaolong, S.Khorram, 1998. The effects of image misregistration on the accuracy of remotely sensed change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 36(5), pp.1566~1577

Gary P, Steve W., 2000. Semautomated linear feature extraction using a knowledge rich object data model. *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXIII Part B3: Amsterdam: pp.740-747

Li Deren, Sui Haigang, Xiao Ping, 2002. Automatic Change Detection of Geo-spatial Data from Imagery, *International Archives of Photogrammetry and Remote Sensing*, Xi'an, China, Vol.XXXIV, Part 2, pp.245-251.

Li Deren, Wang Shuliang, Li Deyi, 2006. *Spatial Data Mining Theories and Applications*, Science Press, Beijing, China

Li Zhilin, Zhu Qing, 2003. *Digital Elevation Model (2nd Edition),* Press of Wuhan University, Wuhan, China.

Sui Haigang, 2002. *Automatic change detection for road-networks base on features*, Ph.D dissertation, Wuhan University. China

Wang Lei, Tang guoan, Liu Xuejun, Long Yi, Wang Chun, 2004. Terrain Complexity Index and Its Derivation from DEMs, *Bullet in of Soil and Water Conservation*, (24)4, pp.55-58

Wang Yongqing, 1998. *Theories and Methods of Artificial Intelligence*, Xi'an Jiaotong University, Xi'an, China

Xia Song, 2006. *Terrain Change Detection and Data Updating with Arial Imagery*, Ph.D dissertation, Wuhan University, China

Xiao Ping, 2001. *Research on Land Use/Cover Change Detection*, Ph.D dissertation, Wuhan University, China

Zhang Lu, 2004. *Change Detection in Remotely Sensed Imagery Using Multivariate Statistical Analysis*, Ph.D Dissertation, Wuhan University, China

Zhang Xiaodong, 2005. *The Theory and Methods of Change Detection Based on Remotely Sensed Imagery and GIS*, Ph.D Dissertation, Wuhan Univesity, China