Spatial and temporal classification of synthetic satellite imagery: land cover mapping and accuracy validation

Yong XU and Bo HUANG*

Department of Geography and Resource Management, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China

(Received 11 August 2013; final version received 26 December 2013)

This study focused on land cover mapping based on synthetic images, especially using the method of spatial and temporal classification as well as the accuracy validation of their results. Our experimental results indicate that the accuracy of land cover map based on synthetic imagery and actual observation has a similar standard compared with actual land cover survey data. These findings facilitate land cover mapping with synthetic data in the area where actual observation is missing. Furthermore, in order to improve the quality of the land cover mapping, this research employed the spatial and temporal Markov random field classification approach. Test results show that overall mapping accuracy can be increased by approximately 5% after applying spatial and temporal classification. This finding contributes towards the achievement of higher quality land cover mapping of areas with missing data by using spatial and temporal information.

Keywords: land cover mapping; synthetic data; spatial and temporal classification

1. Introduction

Given its comparatively large area coverage and high spatial resolution (1–4) with appropriate spectral and temporal resolutions, Landsat satellite data have been the most widely used data for global or regional land cover/mapping applications and resource management (5, 6). However, due to physical and natural constraints, such as undesirable atmospheric effects (i.e. cloud, haze, etc.) and hardware failure, the problems of missing data are prominent for the satellite. To tackle these problems which Landsat 5 and 7 have failed to achieve, Landsat 8 was finally launched on 11 February 2013. Although it has been the longest record of Landsat series to provide scientifically valuable earth observation data, similar to its predecessors, the problem of data missing problem which is affected by the atmospheric conditions is still unsolved.

To provide much more continuous Landsat data with higher quality, a possible way of solving the Landsat gap is to integrate multi-temporal MODIS and Landsat data by using a spatial and temporal adaptive reflectance fusion model (STARFM) to produce more synthetic Landsat-like data at the required times (7–11). Previous studies have proved that STARFM-based synthetic imagery are well suitable for vegetation variation monitoring on a seasonal or inter-annual scale, even in areas with heterogeneous landscapes (8, 11). However, the question of whether these synthetic images are suitable for land cover mapping remains uncertain and their mapping accuracy has not been well verified. Therefore, this study aims to validate the land cover mapping accuracy with STARFM-based synthetic data and investigate the possibility of the improvement of mapping accuracy by using spatial and temporal contextual information.

To date, different land cover classification approaches have been adopted in producing land cover map with satellite data, including the maximum likelihood approaches, support vector machine classification approaches, decision tree classifiers, and neural network approaches (12). In view of the insufficiency of single data source for relevant applications, some data fusion approaches are developed to integrate multi-source satellite data for more effective applications (13): the fusion approach presented by Zhang (13) makes use of both synthetic aperture radar (SAR) and optical image for land cover change detection (14); while Li (15) combined both the remotely sensed images and GIS data for land cover mapping and change detection. These approaches take advantages of spectral and spatial information of satellite data from multi-source satellite data, which are benefit for more accurate land cover mapping and change detection. However, they are only limited to the data fusion with spectral and spatial information. In addition to spectral and spatial information, multi-temporal information is another important characteristic of satellite data, but its land cover mapping applications have rarely been investigated and explored (16). Some scholars did attempt to apply multi-temporal images for generating high-quality land cover maps, achieving some impressive results (16–18). However, despite its wide applicability and potential usages, these multi-temporal-based classification approaches have not been well applied and tested for land cover mapping with synthetic data, not to mention STARFM-based synthetic data which needs investigation. Therefore, this research comparatively studied both the
conventional neural network classification approach and the cutting-edge spatial and temporal random field classification approach for the mapping task with synthetic data.

The rest of this paper is organized as follows. Section 2 presents both the fused approach and spatial and temporal classification approach, which have been used to generate the synthetic data and the land cover map, respectively. Section 3 briefly introduces the study area, and some experimental results are given in Section 4. Finally, Section 5 summarizes this paper.

2. Methodology

2.1. STARFM-based synthetic data

Synthetic data are generated by using STARFM (7). In this model, it is assumed that the surface reflectance of MODIS data for a homogenous area is nearly the same as Landsat reflectance data at the same area, and then their relationship can be modeled as below:

\[ L(x_i, y_j, t_k) = M(x_i, y_j, t_k) + \xi_k \]  

(1)

where \((x_i, y_j)\) means the spatial location of a given point in MODIS and Landsat data; \(t_k\) means the acquisition time; and \(\xi_k\) means systematical error between MODIS and Landsat data, which might belong to their differences in terms of spectral bandwidths, spatial scales, viewing, and solar geometries.

Furthermore, the relationship between MODIS and Landsat data at time \(t_0\) can also be modeled with the following formula:

\[ L(x_i, y_j, t_0) = M(x_i, y_j, t_0) + \xi_0 \]  

(2)

If the systematical error is assumed to remain unchanged over time from \(t_0\) to \(t_k\) at location \((x_i, y_j)\), then \(\xi_k = \xi_0\). Therefore, the following formula can be inferred from the above Equations (1 and 2):

\[ L(x_i, y_j, t_0) = M(x_i, y_j, t_0) + L(x_i, y_j, t_k) - M(x_i, y_j, t_k) \]  

(3)

In order to reduce the systematical error from time \(t_0\) to \(t_k\) and achieve more accurate predicted result, neighborhood information of the predicted point is utilized in a weighted way, as the following weighted operation shows:

\[ L(x_i, y_j, t_0) = \sum_{j=1}^{n} \sum_{k=1}^{n} W_{ijk} \times (M(x_i, y_j, t_0) \]  

\[ + L(x_i, y_j, t_k) - M(x_i, y_j, t_k) \]  

(4)

where \(w\) means the size of the window; \(W_{ijk}\) means the weight coefficient.

The weight for each neighborhoods of the predicted point is decided by three important terms: spatial, temporal, and spectral similarity. Details about the choice of the weight can refer to Ref. (7).

2.2. Spatial and temporal classification

Melgani and Serpico (16) formulated a spatial and temporal classification approach, in which spatial and temporal contextual information is utilized to improve classification accuracy considerably. Since it takes advantages of all the spatial, temporal, and spectral information for land cover classification – some crucial elements in generating high-quality land cover map with multi-temporal and multi-spectral synthetic or actual satellite data – this approach is adopted in this study.

In this approach, multi-temporal images are required for generating land cover map at the prediction time. As shown in Figure 1, three continuous observation fields \(F_t\), \(F_{t-1}\), and \(F_{t+1}\) reflect acquired images at times \(t\), \(t-1\), and \(t+1\), respectively, while \(G_t\), \(G_{t-1}\), and \(G_{t+1}\) represent their label fields. As land cover map (or label field) at the time \(t\), for example, it can be estimated with a maximum a posteriori estimation approach by using multi-temporal images at times \(t\), \(t-1\), and \(t+1\):

\[ g_t = \arg \max_{g_t} P(G_t/F_t, F_{t-1}, G_{t-1}, G_{t+1}) \]  

(5)

where \(G_{t-1}\) and \(G_{t+1}\) are initial label fields at times \(t-1\) and \(t+1\), respectively.

Since it is difficult to maximize the joint probability problem in Equation (5), Besag (19) proposed an iterated conditional modes algorithm which maximizes local conditional probabilities sequentially. Specifically, the right term of Equation (5) can be implemented pixel by pixel with a local approximation instead. As pixel \(i\), for example, its local optimization objective is expressed as below.

\[ P(g_t(i)/F_t, F_{t-1}, G_{t-1}, G_t, G_{t+1}) \approx P(f_t(i)/g_t(i)) \]  

\[ \times P(g_t(i)/C_{sp}) \]  

(6)

where \(g_t(i)\) is the label of pixel \(i\) at time \(t\); \(G_{t-1}\) is a label field of the image at time \(t\) except for pixel \(i\); \(C_{sp}\) reflects
spatial contextual information from images at time \( t \), while \( C_{\text{sp}} \) represents temporal contextual information from images at times \( t-1 \), \( t \), and \( t+1 \).

According to the Hammersley–Clifford theorem \( (20) \), the maximum solution of Equation (6) can be transformed into a local energy minimization problem.

\[
P(f_t(i)/g_t(i)) \propto P(g_t(i)/C_{\text{sp}}, C_p) = \frac{1}{Z} e^{-U_t(g_t(i), f_t(i), C_{\text{sp}}, C_p)}
\]

(7)

where \( U \) can be considered as an energy term and \( Z \) a constant for normalization.

Furthermore, the spatial and temporal contextual information of Equation (7) can be processed separately for simplicity, as shown in below:

\[
U_t(g_t(i), f_t(i), C_{\text{sp}}, C_p) = U_t(g_t(i), f_t(i)) + \theta_{tp} \times U_{\text{sp}}(g_t(i), C_{\text{sp}}) + \theta_{tp} \times U_{\text{tp}}(g_t(i), C_p)
\]

(8)

where \( U_t(g_t(i), f_t(i)) \) is the prior probability energy term, \( U_{\text{sp}}(g_t(i), C_{\text{sp}}) \) is the spatial energy term, \( U_{\text{tp}}(g_t(i), C_p) \) is the temporal energy term, and \( \theta_{tp} \) and \( \theta_{tp} \) are spatial and temporal weighted parameters, respectively.

Specifically, the spatial and temporal energy terms in Equation (8) are estimated as below:

\[
U_{\text{sp}}(g_t(i), C_{\text{sp}}) = -\sum_{g(k) \in C_{\text{sp}}} I(g_t(i), g(k))
\]

(9)

\[
U_{\text{tp}}(g_t(i), C_p) = -\sum_{g(v) \in C_p} P(g_t(i), g(v))
\]

(10)

where \( I(, ,) \) is an indicator function, if \( g_t(i) = g_t(k) \), then 1, otherwise 0; while \( I(, ,) \) is the transition penalty function from the class \( g_t(i) \) to another class \( g(v) \).

The overall procedure of this approach is summarized as below:

1. Acquisition of multi-temporal satellite data, for example, \( F_t, F_{t-1} \), and \( F_{t+1} \) at times \( t, t-1 \), and \( t+1 \), respectively. Each one was classified into \( N \) classes as initial label fields, such as \( G_t, G_{t-1} \), and \( G_{t+1} \).

2. Spatial and temporal parameters of this model are estimated by using the maximum likelihood estimation method \( (21, 22) \). For simplicity, these values can be chosen between 0.5 and 1 according to the study of Melgani and Serpico \( (16) \).

3. Since model parameters are derived from the above two steps, the labels of all pixels are updated one by one according to Equation (8).

3. Case study

3.1. Study area

As shown in Figure 2, a study area in Harvard forest (HARV) of Western Massachusetts, United States, is chosen as a case study. It mainly covers with evergreen forest and deciduous forest, some conifer and mixed hardwood-conifer land covers \( (1) \).

Several sets of satellite data of the Harvard forest are acquired, including both the Landsat and MODIS data from year 2000–2002. In addition, the land cover surveying data for this area for year 2000, 2001, and 2002 are also collected from a data validation group called Bigfoot \( (1) \).

3.2. Synthetic data generation

In the following test, Landsat data for year 2001 are assumed to be unknown and the actual observation for this year is used as validation data. To predict satellite data for 2001, two pairs of high-resolution Landsat and low-resolution MODIS images for 25 August 2000 and 22 July 2002, as well as a low-resolution MODIS image for 5 September 2001, are applied to generate a high-resolution synthetic reflectance image for 5 September 2001. Figure 3(b) shows the predicted reflectance image for 5 September 2001 by using the STARFM model.

In the following, synthetic STARFM-based Landsat data for 2001, with the actual Landsat data for 2000, 2001, and 2002, are applied and validated for land cover mapping applications in this study area.

---

Figure 2. Study area at Harvard Forest, US.
Figure 3. MODIS data and the predicted synthetic Landsat data for 5 September 2001 for the HARV study area.

(a) MODIS data

(b) Synthetic data with STARFM

Figure 4. Landsat or synthetic data and the generated land cover map with neural network classification method for the HARV study area.

(a) Actual Landsat data for September 5, 2001

(b) Land cover map from actual data

(c) Synthetic data

(d) Land cover map from synthetic data
4. Validation of land cover mapping based on synthetic reflectance data

4.1. Land cover mapping with conventional classification

The study area mainly has six different types of land cover, including water, grassland, wetland, deciduous forest, evergreen forest, and built-up area. Conventional neural network classification approach could be applied to produce a land cover map with six classes. After this approach was applied to the actual observation as shown in Figure 4(a) for 5 September 2001, a land cover map was generated (see Figure 4(b)). Similarly, the synthetic reflectance image shown in Figure 4(c) for 5 September 2001 can also be used to generate a land cover map (see Figure 4(d)).

Two accuracy assessment indexes, including the overall accuracy (OA) and Kappa coefficient, were used by comparing the mapping result from synthetic data with the actual land cover surveying map for 2001. After compared with the actual land cover map for 2001, the accuracy statistics of the generated land cover map with neural network classification method by using synthetic or actual data are given in the left part of Table 1. These results indicated that the accuracy of land cover map based on synthetic imagery and actual observation has a similar standard, nevertheless the quality of land cover map from synthetic data is slightly worse than the one from actual observation.

4.2. Land cover mapping with spatial and temporal classification

The spatial and temporal classification approach introduced in Section 2.2 is applied to validate the classification accuracy by using synthetic data. Both the actual observations for 2000 and 2002, as well as the synthetic data for 5 September 2001, are used to generate the land cover map for 5 September 2001 at the HARV study area, since the actual observations for 25 August 2000 and 22 July 2002 are closest to the prediction date. After the spatial and temporal classification was applied, Figure 5(a) shows the land cover map based on images for 2000 and 2002, as well as the synthetic data for 5 September 2001, while Figure 5(b) shows the land cover map by using actual observations from 2000 to 2002.

4.3. Results and discussion

The accuracy statistics for mapping results of both neural network and spatial and temporal classification approaches are given in Table 1, from which there are two main findings. First, the mapping accuracies based on synthetic data and actual data are nearly the same, regardless of which method applied. For the results with spatial and temporal classification in 2001, the overall mapping accuracy based on synthetic data is 81.19%, while the true overall mapping accuracy with actual observation is 81.63%.

Second, following the application of the spatial and temporal classification method, the mapping accuracy improved around 5% than the conventional classification

| Data (2001)      | Neural network | Spatial-temporal classification |
|------------------|----------------|---------------------------------|
|                  | OA (%)         | Kappa                           | OA (%) | Kappa |
| Synthetic data   | 75.67          | 0.4632                          | 81.19  | 0.5322|
| Actual data      | 76.72          | 0.4880                          | 81.63  | 0.5491|

Note: actual data for 2000 and 2002 are used for spatial-temporal classification.

Figure 5. Generated land cover map with spatial-temporal classification method by using synthetic or actual data for the HARV study area. (a) Land cover map from synthetic data using spatial-temporal classification method; (b) Land cover map from actual data using spatial-temporal classification method.
approach. For the classification accuracy with synthetic data, its OA increases from 75.67% with the neural network classification method to 81.19% with the spatial and temporal classification for this study area.

5. Summary
STARFM-based synthetic data have been proved to be quite suitable for vegetation monitoring and variation analysis, but rarely investigated for the applications of land cover mapping. In this paper, several tests were designed to verify whether synthetic images are suitable for land cover mapping. The result of our tests includes three aspects. First, according to our experiment, we find that the spatial and temporal reflectance fusion model is suitable for fusing images from nearly the same season of different years. Second, after the synthetic data were used to generate a land cover map using the neural network classification method, the experimental results show that the overall mapping accuracy by using synthetic data is nearly the same as that by using actual observation for HARV study area. The OA is 75.67% vs. 76.72 (synthetic data vs. actual data) for the HARV study area for 11 November 2001. This means that the synthetic data can be used to generate a land cover map; however, it is less accurate than actual observation. Finally, multi-temporal image classification was applied to both the synthetic reflectance data and its subsequent images. Experimental results show that the mapping accuracy by using multi-temporal images is higher than that by using a single image. Most importantly, the degree of mapping accuracy with synthetic data in the multi-temporal image classification model is nearly the same as that when actual observation was used in the model. This means that the synthetic data can be used for generating land cover maps whenever the actual high-resolution images are lost, and its degree of accuracy, which is similar to that with actual observation, be the result obtained from the general classification method or a spatial-temporal classification method, such as the spatial and temporal Markov random field method.

In addition, the spatial-temporal classification method is useful for generating more high-quality land cover maps, as our tests show that there is around 5% improvement when no less than three images are used for these study areas.

Notes on Contributors
Yong Xu is currently a research assistant in the Department of Geography and Resource Management at the Chinese University of Hong Kong, Hong Kong. His current research interest includes image processing.

Bo Huang is currently a professor in the Department of Geography and Resource Management at the Chinese University of Hong Kong, Hong Kong. His research interests include remote sensing and GIS.

Acknowledgments
This work was supported in part by the National High-Tech R&D Program (863 program) under grant number 2009AA122004, the National Natural Science Foundation of China under grant number 60171009, and the Hong Kong Research Grant Council under grant number CUHK 444612.

References
(1) Cohen, W.B.; Maiersperger, T.K.; Yang, Z.; Gower, S.T.; Turner, D.P.; Ritts, W.D.; Berterretche, M.; Running, S.W. Comparisons of Land Cover and LAI Estimates Derived from ETM+ and MODIS for Four Sites in North America: A Quality Assessment of 2000/2001 Provisional MODIS Products. Remote Sens. Environ. 2003, 88 (3), 233–255.
(2) Cohen, W.B.; Goward, S.N. Landsat’s Role in Ecological Applications of Remote Sensing. BioScience 2004, 54 (6), 535–545.
(3) Goward, S.N.; Masek, J.G.; Williams, D.L.; Irons, J.R.; Thompson, R.J. The Landsat 7 Mission: Terrestrial Research and Applications for the 21st century. Remote Sens. Environ. 2001, 78, 3–12.
(4) Wulder, M.A.; White, J.C.; Goward, S.N.; Masek, J.G.; Irons, J.R.; Herold, M.; Cohen, W.B.; Loveland, T.R.; Woodcock, C.E. Landsat Continuity: Issues and Opportunities for Land Cover Monitoring. Remote Sens. Environ. 2008, 112 (3), 955–969.
(5) Knorn, J.; Rabe, A.; Radloff, V.C.; Kuenzer, T.; Kozak, J.; Hostert, P. Land Cover Mapping of Large Areas Using Chain Classification of Neighboring Landsat Satellite Images. Remote Sens. Environ. 2009, 113 (5), 957–964.
(6) Masek, J.G.; Huang, C.; Wolfe, R.; Cohen, W.; Hall, F.; Kutler, J.; Nelson, P. North American Forest Disturbance Mapped from a Decadal Landsat Record. Remote Sens. Environ. 2008, 112, 2914–2926.
(7) Gao, F.; Masek, J.; Schwaller, M.; Hall, F. On the Blending of the Landsat and MODIS Surface Reflectance Predicting Daily Landsat Surface Reflectance. IEEE Trans. Geosci. Remote Sens. 2006, 44 (8), 2207–2218.
(8) Hilker, T.; Wulder, M.A.; Coops, N.C.; Linke, J.; McDermid, G.; Masek, J.G.; Gao, F.; White, J.C. A New Data Fusion Model for High Spatial- and Temporal-resolution Mapping of Forest Disturbance Based on Landsat and MODIS. Remote Sens. Environ. 2009, 113 (8), 1613–1627.
(9) Hilker, T.; Wulder, M.A.; Coops, N.C.; Seitz, N.; White, J.C.; Gao, F.; Masek, J.G.; Stenhouse, G. Generation of Dense Time Series Synthetic Landsat Data Through Data Blending with MODIS Using a Spatial and Temporal Adaptive Reflectance Fusion Model. Remote Sens. Environ. 2009, 113 (9), 1988–1999.
(10) Roy, D.P.; Ju, J.; Lewis, P.; Schaaf, C.; Gao, F.; Hansen, M.; Lindquist, E. Multi-temporal MODIS-Landsat Data Fusion for Relative Radiometric Normalization, Gap Filling, and Prediction of Landsat Data. Remote Sens. Environ. 2008, 112, 3112–3130.
(11) Zhu, X.; Chen, J.; Gao, F.; Chen, X.; Masek, J.G. An enhanced Spatial and Temporal Adaptive Reflectance Fusion Model for Complex Heterogeneous Regions. Remote Sens. Environ. 2010, 114 (1), 2610–2623.
(12) Huang, C.; Davis, L.S.; Townshend, J.R.G. An Assessment of Support Vector Machines for Land Cover Classification. Int. J. Remote Sens. 2002, 23 (4), 725–749.
(13) Zhang, J.X. Multi-source Remote Sensing Data Fusion: Status and Trends. Int. J. Image Data Fusion 2010, 1 (1), 5–24.
(14) Zeng, Y.; Zhang, J.X.; van Genderen, J.L.; Zhang, Y. Image Fusion for Land Cover Change Detection. *Int. J. Image Data Fusion*. 2010, 1 (2), 193–215.

(15) Li, D.R. Remotely Sensed Images and GIS Data Fusion for Automatic Change Detection. *Int. J. Image Data Fusion* 2010, 1 (1), 99–108.

(16) Melgani, F.; Serpico, S.B. A Markov Random Field Approach to Spatio-temporal Contextual Image Classification. *IEEE Trans. Geosci. Remote Sens.* 2003, 41 (11), 2478–2487.

(17) Shimabukuro, Y.E.; Duarte, V.; Arai, E.; Freitas, R.M.; Martini, P.R.; Lima, A. Monitoring Land Cover in Acre State, Western Brazilian Amazonia, Using Multitemporal Remote Sensing Data. *Int. J. Image Data Fusion* 2010, 1 (4), 325–335.

(18) Kandrika, S.; Ravisankar, T. Multi-temporal Satellite Imagery and Data Fusion for Improved Land Cover Information Extraction. *Int. J. Image Data Fusion* 2011, 2 (1), 61–73.

(19) Besag, J. On the Statistical Analysis of Dirty Pictures. *J. Royal Stat. Soc., Ser. B* 1986, 48, 259–302.

(20) Hammersley, J.M.; Clifford, P. Markov Fields on Finite Graphs and Lattices. Unpublished paper, 1971. http://www.statslab.cam.ac.uk/~grg/books/hammfest/hamm-cliff.pdf (accessed Feb 4, 2014).

(21) Besag, J. Spatial Interaction and the Statistical Analysis of Lattice Systems. *J. Royal Stat. Soc.* 1974, B36, 192–236.

(22) Li, S.Z. *Markov Random Field in Computer Vision*; Springer-Verlag: Berlin, 1995.