SIMULATING REMOTELY SENSED IMAGERY FOR CLASSIFICATION EVALUATION

Desheng Liu

Department of Geography and Department of Statistics, The Ohio State University
1036 Derby Hall, 154 N Oval Mall, Columbus, OH, 43210, USA (liu.738@osu.edu)

KEY WORDS: Simulation, Remote Sensing, Image Classification, Accuracy Assessment, Simulated Annealing

ABSTRACT:

The remote sensing community has long been active in developing, evaluating, and comparing different classification algorithms using a variety of remotely sensed imagery. As an integral component of image classification, accuracy assessment is usually conducted to evaluate the agreement between the classified map and the corresponding reference data. However, current accuracy assessment practices are limited by the difficulties in obtaining high quality reference data and the lack of spatial representation of classification uncertainties. To overcome these limitations, we developed a simulation approach to obtaining the desired reference data for better evaluation of classification algorithms. The simulation approach involves three components: 1) a real image scene, 2) a reference map, and 3) a simulated image scene. The real image scene is assumed as a random realization of a spectral probability model governed by an unknown underlying process, which is defined as the ground truth of the classification of the image scene. The reference map represents a reasonable estimate of the unknown process that generates the real image scene. The simulated image scene is generated as a random realization of the spectral probability model governed by the estimated process represented in the reference map. Specifically, an initial simulated image is firstly generated by independently sampling based on probability distributions estimated from the real image scene and reference map. Then, the initial simulated image is iteratively perturbed using simulated annealing to create a final simulated image which has similar spectral, spatial, and textual properties to the real image scene. The simulation approach was applied to a Landsat TM image scene and promising results were achieved.

1. INTRODUCTION

Image classification is one of the most fundamental applications of remotely sensed data (Jensen 2004). Thematic maps generated from the classification of remotely sensed imagery are widely used in many environmental, ecological, and social-economic studies. Driven by the need for better classification results, the remote sensing community has long been active in developing, evaluating, and comparing different classification algorithms (Erbek et al, 2004; Flygare, 1997; Liu et al, 2006; Lu et al, 2004). The past decades have witnessed the developments of a large number of computer-based classification algorithms using a variety of remotely sensed imagery across different fields (Landgrebe, 2003; Lu and Weng, 2007; Tso and Mather, 2001). With the increased availability and enhanced capability of remotely sensed data, the efforts in advancing classification algorithms are expected to continue.

As an integral component of image classification, accuracy assessment (Congalton and Green, 1999; Foody, 2002) is usually conducted to evaluate the agreement between the classified map and the corresponding reference data (or ground truth). For map users, the accuracy assessment serves as an indicator on the accuracy of the classified map. For classification algorithm developers, the accuracy assessment is an essential procedure to evaluate the performances of classification algorithms. Current practices on accuracy assessment are primarily based on various global accuracy indices (e.g. overall accuracy, user's accuracy, producer's accuracy, and kappa coefficient) derived from a confusion or error matrix, which is built upon the reference data for a small subset of the classified imagery. The reference data are usually obtained from field survey, visual interpretation of high spatial resolution imagery, or other existing maps.

However, there are two major limitations for the current accuracy assessment methods. The first limitation is related to the quality of reference data. One important requirement for the accuracy indices to be valid estimators of the true map accuracy is that the reference data are representative to the whole imagery and accurately reflect the ground truth (Congalton and Green, 1999; Foody, 2002). Representative reference data demand a careful design of probability sampling. In field survey, rigorous probability sampling design is often hard to implement due to physical constraints or high costs. The interpretation of high spatial resolution imagery can be biased by the tendency for selecting homogeneous regions and avoiding boundaries and complex regions. Existing maps often contain classification errors and are outdated compared to the classified image. Moreover, all the above three means of obtaining reference data are subject to registration errors of the reference data to the classified imagery. Consequently, accuracy assessment based on non-representative, inaccurate, and mis-registered reference data is not reliable for both map users and classification algorithm developers (Congalton and Green, 1999; Foody, 2002). The second limitation lies in the lack of spatial representation of classification uncertainties in the current accuracy assessment methods. It is well-known that classification errors are not randomly distributed across the imagery but have strong spatial patterns. Knowledge on the spatial patterns of classification errors can inform map users on the classification uncertainty and possible error propagation. For classification algorithm developers, a good understanding on the error patterns across the imagery helps to better evaluate the spatial aspects of classification algorithms and provide insights for further improvements. This is particularly true when contextual and object-based classification algorithms are assessed. Unfortunately, the accuracy indices derived from confusion matrix only give a global view of classification

results but provide no information on the spatial distribution of classification errors.

The two limitations discussed above call for more comprehensive and accurate reference data with which a more reliable and complete accuracy assessment can be performed. For map users, this remains an open question as it depends on the real situation of the specific mapping area. For classification algorithm developers, however, the reference data are not necessarily limited to the real situation given that the main objective is to evaluate classification algorithms. Therefore, we propose to use image simulation as a new approach to obtaining the desired reference data for better evaluation of classification algorithms. The purpose of this paper is to develop a new approach to simulating remotely sensed imagery for fully evaluating classification algorithms. The rest of the paper is organized as follows. We first introduce our conceptual model for simulating remotely sensed imagery followed by detailing the procedures for image simulation. Then we present the results using Landsat TM image. Finally, we discuss issues related to image simulation and conclude with a summary.

2. METHODS

2.1 The Conceptual Model

As illustrated in Figure 1, the conceptual model used for image simulation consists of three components: 1) a real image scene, 2) a reference map, and 3) a simulated image scene. The real image scene can be any remotely sensed imagery of interest. It provides basic information on the spectral, spatial, and textual aspects of a real remotely sensed image scene. The real image scene is assumed to be a random realization of spectral reflectance governed by an unknown underlying process, which is defined as the ground truth for the classification of the image scene. The reference map represents a reasonable estimate of the unknown process that generates the real image scene. As an estimate, the reference map may not accurately represent the ground truth for the classification of the real image scene. Hence, accuracy assessment based on the reference map may not be reliable. However, the reference map serves as a connection between the real image scene and the simulated image scene. The simulated image scene is generated as a random realization of the spectral reflectance, which has similar spectral, spatial, and textual properties to the real image scene, governed by the estimated process represented in the reference map. Therefore, the reference map is the true representation of the classification of the simulated image scene.
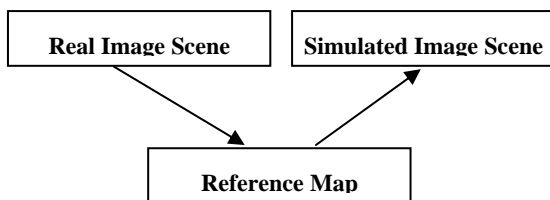

Figure 1. The conceptual model of image simulation

Based on the conceptual model, the classification algorithm developers can apply any classification algorithm to the simulated image scene and evaluate its performance by accuracy assessment based on the reference map. This simulation approach to accuracy assessment has three

advantages over the traditional approaches using limited reference data of the real image scene. Firstly, the reference map is free from classification and registration errors to the simulated image scene. Secondly, the reference map consists of the entire sample space, so it allows classification developers to experiment with different sampling designs. Finally, the reference map allows the spatial representation of classification errors by evaluating every classified pixel.

2.2 Reference Map Generation

One approach to generating a reference map is through supervised classification of the real image scene based on the real reference data collected in a traditional way. The classified image can be further post-processed (e.g. removing "salt-and-pepper" effects) to entail certain cartographic generalization and achieve a map-like quality. Nevertheless, the final reference map may still have certain level of classification errors due to the imperfection of the reference data and the classifier, but it should approximately reflect the general pattern of the underlying process that generates the real image scene. Since the main goal at this stage is not to test the classification algorithms but to have a baseline reference map for simulating images, the final reference map will be used as a starting point for the following image simulation.

2.3 Image Simulation

The main goal of the image simulation is to generate a simulated image scene so that 1) its underlying process is governed by the reference map, and 2) its spectral, spatial, and textual property is similar to that of the real image scene. In doing so, we utilize an optimization algorithm called simulated annealing for the simulation. The basic idea of simulated annealing is to perturb an initial (or seed) image according to a suitable annealing schedule in order to create a desired (optimal) image which minimizes a set of user-defined objective functions (Geman and Geman, 1984; Goovaerts, 1997; Burnicki et al, 2007). In what follows, we describe the generation of initial image, define objective functions, and specify perturbation mechanism.

2.3.1 Initial Image Generation: Based on the real image scene and the reference map, the probability distribution of spectral reflectance is estimated for each class on the reference map. Gaussian models can be used for simple distributions while Gaussian mixture models may be needed for more complex distributions. Denote the initial image scene by $\left\{ X^{(0)}(u_1),...,X^{(0)}(u_N) \right\}$, where u_i is the spatial location of pixel i $(i = 1,...,N)$. The initial image scene is then generated as a random realization of the estimated probability distribution models. Specifically, for each pixel, a random sample is drawn from the probability distribution conditional on its class type on the reference map. In this way, the spectral properties of the real image scene were captured in the simulated image. However, the initial image scene is purely a collection of independently, identically distributed (I.I.D.) samples of each class. The following perturbation will make it a more realistic image scene for classification analysis.

2.3.2 Objective Functions: The initial image does not have any spatial and textual patterns that a real image scene should have. The purpose of the perturbation is to modify the initial image to achieve the desired spatial and textual properties. To do so, objective functions are needed to guide the perturbation.

Therefore, we define two objective functions to represent the spatial and textual components of the real image scene. Since a single image band is often used to model the spatial and textual properties of a real image scene, we construct the two objective functions based on grayscale imagery in the following. For multispectral imagery, a principle component analysis (PCA) can be applied firstly; and the following objective functions are then calculated using the major principle component.

The first objective function, denoted by O_1, characterizes the spatial autocorrelation of an image scene using a semivariogram model. O_1 is calculated as the root mean squared difference between the semivariances of the real image scene and those of the simulated image scene over a set of lags:

$$O_1 = \sqrt{\frac{1}{L}\sum_{l=1}^{L}\left(\gamma_S(l) - \gamma_R(l)\right)^2} , \qquad (1)$$

where $l = 1,...,L$ are the spatial lags; $\gamma_S(l)$ and $\gamma_R(l)$ are the empirical semivariances at lag l for the simulated image scene and the real image scene respectively.

The second objective function, denoted by O_2, quantifies the textual pattern of an image scene using a texture measure. Specifically, the textual measure is defined by the local entropy of a 9×9 image window centred on each pixel. O_2 is calculated as the root mean squared difference between the pixel-wise entropy measures of the real image scene and the simulated image scene:

$$O_2 = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(E_S(u_i) - E_R(u_i)\right)^2} , \qquad (2)$$

where u_i is the spatial location of pixel i $(i = 1,...,N)$; $E_S(u_i)$ and $E_R(u_i)$ are the entropy measures at pixel i for the simulated image scene and the real image scene respectively. The two objective functions are then combined as one overall objective function using a suitable weight ω:

$$O = O_1 + \omega O_2 . \qquad (3)$$

2.3.3 Iterative Perturbation: The initial overall objective function $O^{(0)}$ is firstly calculated based on the initial simulated image. A sequence of iterative stochastic perturbation $(m = 1,...,M)$ is then introduced to modify the initial image to minimize the overall objective function. Denote the simulated image scene after the m-th iteration as $\left\{X^{(m)}(u_1),...,X^{(m)}(u_N)\right\}$.

At the m-th iteration, one perturbation is initiated by swapping a pair of randomly chosen pixels at u_i and u_j. That is, $X^{(m)}(u_i)$ and $X^{(m)}(u_j)$ are updated as follows

$$\begin{cases} X^{(m)}(u_1) = X^{(m-1)}(u_2) \\ X^{(m)}(u_2) = X^{(m-1)}(u_1) \end{cases} . \qquad (4)$$

The m-th overall objective function $O^{(m)}$ is then calculated based on $\left\{X^{(m)}(u_1),...,X^{(m)}(u_N)\right\}$. Finally, a decision is made on whether to accept the m-th perturbation according the following rules:

1. When $O^{(m)} \leq O^{(m-1)}$, accept the m-th perturbation;

2. When $O^{(m)} > O^{(m-1)}$, accept the m-th perturbation with a probability $p = \exp(-m/\lambda)$, where λ is a constant to control the rate of the decrease of the probability p with respect to the increase of m.

If the decision is to reject the m-th perturbation, then the update in (4) is not taken. Equivalently, $\left\{X^{(m)}(u_1),...,X^{(m)}(u_N)\right\}$ is the same as $\left\{X^{(m-1)}(u_1),...,X^{(m-1)}(u_N)\right\}$.

The above iteration process proceeds until a predefined objective function value is achieved or the maximum number of iterations is reached.

3. RESULTS

We tested our algorithms based on a scene of Landsat TM image. This image scene has a size of 256 by 256 pixels, which are of 28.5 meter resolution, with the exception of the thermal infrared band of 57 meter resolution. For the purpose of image simulation, three bands (band 3, band 5, and band 7) were selected for this study. A false colour representation of the real image scene is shown in Figure 2(a). Two classes (forest and bare) were considered for the reference map in the simulation. The reference map shown in Figure 2(b) was generated by unsupervised classification followed by some post-processing. As introduced before, the reference map is not perfect ground truth to the real image scene. However, it reasonably estimated the underlying process that generated the real image scene, which was used to generate the simulated image scene.
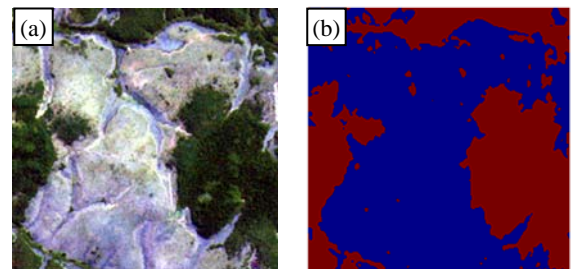

Figure 2. (a) The real image scene, (b) the reference map

Preliminary exploratory analysis on the histograms of spectral data of two classes indicated that multivariate Gaussian models were sufficient for estimating the probability distributions. Therefore, the initial simulated image shown in Figure 3(a) was generated by randomly drawing independent samples from the two multivariate Gaussian models which were estimated from

the real image scene. Figure 3(a) showed that the initial simulated image was extremely noisy due to the lack of spatial and textural patterns.
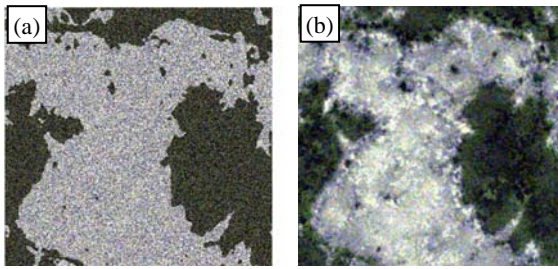

Figure 3. (a) The initial simulated image scene, (b) the final simulated image scene

The differences in spatial and textual patterns between the real image scene and the initial simulated image scene are further illustrated in their local entropy measures in Figure 4(a) and Figure 4(b) and semivariograms in Figure 5. From Figure 4(a) and (b), the local entropy measure of the initial simulated image scene was quite different from that of the real image scene. Figure 5 showed that the semivariance of the initial simulated image scene (dotted blue line) was larger than that of the real image scene (solid red line) at all lags, particularly at the small lags, indicating large nugget effects for the initial simulated image scene.
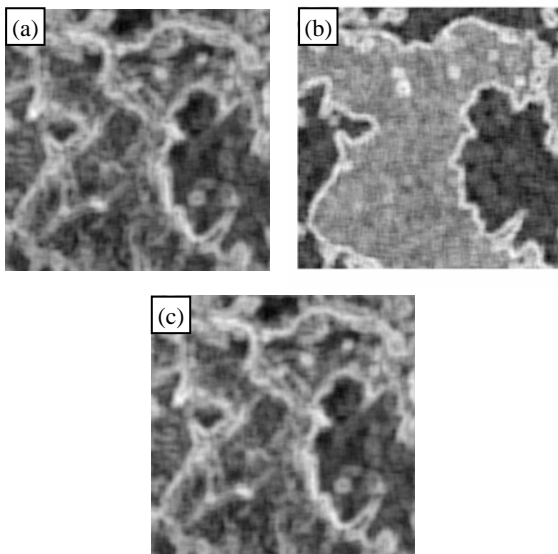

Figure 4. The local entropy measures of (a) the real image scene, (b) the initial simulated image scene, and (c) the final simulated image scene
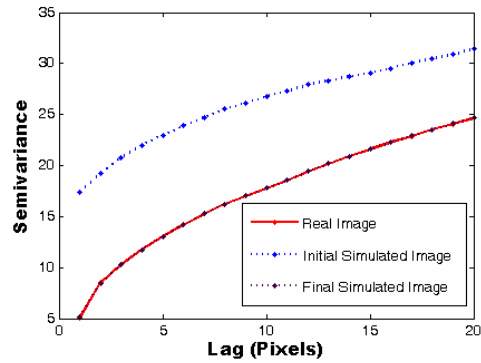

Figure 5. The semivariograms of the real image scene (solid red line), initial simulated image scene (dotted green line), and the final simulated image (dotted black line)

After about one million perturbations using simulated annealing, the final simulated image shown in Figure 3(b) was generated with a pre-defined small threshold value on the overall objective function achieved. Compared with the initial simulated image scene in Figure 3(a), the final simulated image scene showed strong spatial and textual patterns. More importantly, the spatial and textual patterns in the final simulated image scene appeared very similar to those observed in the real image scene. This was further confirmed when comparing their local entropy measures and semivariograms. In Figure 5, the red solid line nearly coincided with the black dotted line, indicating that the final simulated image scene achieved the targeted semivariance of the real image scene at all lags. The local entropy image illustrated in Figure 4(c) also matched perfectly with that in Figure 4(a). Visual comparison of the textures between Figure 2(a) and Figure 3(b) further confirmed this.

4. DISCUSSION

The simulation approach showed promising results for the Landsat TM image scene as evidenced by the desired spatial and textual properties on the final simulated image scene. However, the final simulated image scene differed from the real image scene in some local structural patterns. For example, the linear features in the real image scene were not captured in the final simulated image. This was due to the fact that the two objective functions defined in this paper could not uniquely determine the spatial and textual properties of an image scene. For the purposes of evaluating classification algorithms, this may have the advantage that more simulations can be implemented to explore different spatial and textual patterns.

This study was focused on a relatively small image scene because the use of simulated annealing in the simulation process is computationally expensive for large image scene. As each perturbation only involves two pixels, the changes of objective functions are only affected by the local neighbourhood of the two pixels. Hence, some implementation tips suggested by Goovaerts (1997) can be used to improve the computation efficiency.

The proposed simulation framework is pretty general and should be applicable to any image scenes. This study only tested a relatively simple landscape, where Gaussian models were sufficient as probability distributions. Future research should explore a variety of image scenes with different

complexities in landscape and probability distributions. More sophistic probability models and objective functions may be needed for more complex image scenes.

5. CONCLUSIONS

Current accuracy assessment practices for classification evaluation are limited by the difficulties in obtaining high quality reference data and the lack of spatial representation of classification uncertainties. In this paper, we proposed a new simulation approach to generating remotely sensed imagery in an attempt to obtain ideal reference data for better evaluation of classification algorithms. The simulation approach involves three components: 1) a real image scene, 2) a reference map, and 3) a simulated image scene. Specifically, an initial simulated image is firstly generated by independently sampling probability distributions estimated from the real image scene and reference map. Then, the initial simulated image is perturbed using simulated annealing to create a final simulated image which has similar spectral, spatial, and textual properties of the real image scene. When the reference map is used for accuracy assessment of the classification of the simulated image scene, three advantages can be found: 1) the reference map is free from classification and registration errors to the simulated image scene, 2) the reference map consists of the entire sample space, so it allows classification developers to experiment with many different sampling schemes, and 3) it allows the spatial representation of classification uncertainties.

REFERENCES

Burnicki, A.C., D.G. Brown, and P. Goovaerts, 2007. Simulating error propagation in land-cover change analysis: The implications of temporal dependence. *Computers, Environment and Urban Systems*, 31(3), pp. 282-302.

Congalton, R. G., and K. Green, 1999. *Assessing the accuracy of remotely sensed data: principles and practices*, Lewis Publishers, Boca Raton.

Erbek, F.S., C. Ozkan and M. Taberber, 2004. Comparison of maximum likelihood classification method with supervised artificial neural network algorithms for land use activities, *International Journal of Remote Sensing*, 25, pp. 1733–1748.

Flygare, A-M., 1997. A comparison of contextual classification methods using Landsat TM, *International Journal of Remote Sensing*, 18, pp. 3835–3842.

Foody, A.M., 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80, pp. 185-201.

Geman, S. and D. Geman, 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transactions on. Pattern Analysis and Machine Intelligence,* 6, pp. 721-741.

Goovaerts, P., 1997. *Geostatistics for natural resources evaluation,* Oxford University Press, New York.

Jensen, J.R., 2004. *Introduction to Digital Image Processing: A remote sensing perspective*, 3rd *edition*. Prentice Hall, Piscataway.

Landgrebe, D.A., 2003. *Signal Theory Methods in Multispectral Remote Sensing,* John Wiley and Sons, Hoboken.

Liu, D., M. Kelly, and P. Gong, 2006. A spatial-temporal approach to monitoring forest disease spread using multi-temporal high spatial resolution imagery, *Remote Sensing of Environment*, 101(2): 167-180.

Lu, D., P. Mausel, M. Batistella, and E. Moran, 2004. Comparison of land-cover classification methods in the Brazilian Amazon Basin, *Photogrammetric Engineering and Remote Sensing*, 70, pp. 723–731.

Lu, D. and Q. Weng, 2007. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28(5), pp. 823-870.

Tso, B. and P.M. Mather, 2001. *Classification Methods for Remotely Sensed Data*. Taylor and Francis, New York.