

A SURVEY OF REMOTE SENSING IMAGE CLASSIFICATION APPROACHES

Li Li^a, Jianqiang Gao^{b*} and Balan Sethuramalingam^c

^aSchool of Internet of Things Engineering, Jiangnan University, Wuxi 214122,
P. R. China

^bCollege of Computer and Information, Hohai University, Nanjing 210098,
P. R. China

^cGovernment Arts College, Bharathiar University, Coimbatore 641018,
Tamilnadu, India

Abstract

Remote-sensing image classification is a complex process that may be affected via lots of factors. This paper mainly examines current practices, problems, and prospects of remote-sensing image classification. The aim is to focus on the summarization of major advanced classification methods and the techniques used for improving classification accuracy. This paper review suggests that designing a suitable image processing procedure is a prerequisite for a successful classification of remotely sensed data and the selection of a suitable classification accuracy. Neural network, decision tree classifier, and knowledge-based classification have increasingly become important methods for remote-sensing image classification. However, more research is needed to identify and reduce uncertainties in the image processing chain to improve classification accuracy.

Keywords: remote-sensing image, classification approach, preprocessing of image, feature extraction and selection.

*Corresponding author.

E-mail address: jianqianggao@126.com (*Jianqiang Gao), lili880827@126.com (Li Li)
and balan.sethuramalingam@gmail.com (Balan Sethuramalingam).

1. Introduction

Remote-sensing research focusing on image classification has long attracted the attention of the remote-sensing community because classification results are the basis for many environmental and socioeconomic applications. Many scientists have made great efforts in developing advanced classification methods and techniques for improving classification accuracy [1-9]. However, classifying remotely sensed data into a thematic map remains a challenge because many factors, such as the complexity of the landscape in a study area, selected remotely sensed data, and image-processing and classification approaches, may affect the success of a classification. Although much previous research are specifically concerned with image classification [10, 11]. A review of new methods will be highly valuable for guiding or selecting a suitable classification procedure for a specific study.

The focus of this paper is on providing a summarization of major advanced classification approaches and techniques used for improving classification accuracy, and on discussing important issues affecting the success of image classifications.

2. Classification Process

Remote-sensing classification is a complex process and requires consideration of many factors. The major steps of image classification may include determination of a suitable classification system, selection of training samples, feature extraction, selection of suitable classification methods, and accuracy assessment. The user's need, scale of the study area, economic condition, and analyst's skills are important factors influencing the selection of remotely sensed data, the design of the classification procedure, and the quality of the classification results. The section focuses on the description of the major steps that may be involved in image classification.

2.1. Selection of a classification system and training samples

A suitable classification system and a sufficient number of training samples are prerequisites for a successful classification. Generally speaking, a classification system is designed based on the user's need, spatial resolution of selected remotely sensed data, image-processing and classification algorithms available, and time constraints. In many cases, a hierarchical classification system is adopted to take different conditions into account.

A sufficient number of training samples and their representativeness are critical for image classifications [11-14]. Training samples are usually collected from fieldwork, or from fine spatial resolution aerial photographs and satellite images. Different collection strategies, such as single pixel, seed, and polygon, may be used, but they would influence classification results, especially for classifications with fine spatial resolution image data [13]. When the landscape of a study area is complex and heterogeneous, selecting sufficient training samples becomes difficult. This problem would be complicated if medium or coarse spatial resolution data are used for classification, because a large volume of mixed pixels may occur. So, the selection of training samples must consider the spatial resolution of the remote-sensing data being used, availability of ground reference data, and the complexity of landscapes in the study area.

2.2. Feature extraction and selection

Selecting suitable variables is a critical step for successfully implementing an image classification. Many potential variables may be used in image classification, including spectral signatures, textural or contextual information, and ancillary data. Due to different capabilities in land-cover separability, the use of too many variables in a classification procedure may decrease classification accuracy [15, 16]. It is important to select only the variables that are most useful for separating land-cover or vegetation classes, especially when hyperspectral data are employed. Many approaches, such as principal component analysis, minimum noise fraction transform, discriminant analysis,

decision boundary feature extraction, non-parametric weighted feature extraction, wavelet transform, and spectral mixture analysis [11, 17-23] may be used for feature extraction, in order to reduce the data redundancy inherent in remotely sensed data or to extract specific land-cover information.

Optimal selection of spectral bands for classifications has been extensively discussed in previous reports. Graphic analysis and statistical methods have been used to identify an optimal subset of bands.

2.3. Selection of classification method

Many factors, such as spatial resolution of the remotely sensed data, different sources of data, a classification system, and availability of classification software must be taken into account when selecting a classification method for use. Different classification methods have their own merits. The question of which classification approach is suitable for a specific study is not easy to answer. Different classification results may be obtained depending on the classifier(s) chosen.

2.4. Evaluation of classification performance

Evaluation of classification results is an important process in the classification procedure. Different approaches may be employed, ranging from a qualitative evaluation based on expert knowledge to a quantitative accuracy assessment based on sampling strategies. To evaluate the performance of a classification method, Cihlar et al. [24] proposed six criteria: accuracy, reproducibility, robustness, ability to fully use the information content of the data, uniform applicability, and objectiveness. In reality, no classification algorithm can satisfy all these requirements nor be applicable to all studies, due to different environmental settings and datasets used. However, classification accuracy assessment is the most common approach for an evaluation of classification performance. The error matrix approach is the one most widely used in accuracy assessment [25]. In order to properly generate an error matrix, one must consider the following factors: (1) reference data

collection; (2) classification scheme; (3) sampling scheme; (4) spatial autocorrelation; and (5) sample size and sample unit. The Kappa coefficient is a measure of overall statistical agreement of an error matrix, which takes non-diagonal elements into account. Kappa analysis is recognized as a powerful method for analysing a single error matrix and for comparing the differences between various error matrices.

In summary, the error matrix approach is the most common accuracy assessment approach for categorical classes. Uncertainty and confidence analysis of classification results has gained some attention recently, and spatially explicit data on mapping confidence are regarded as an important aspect in effectively employing classification results for decision making [26, 27].

3. Classification Methods

Recently, lots of classification methods, such as artificial neural networks, fuzzy-sets, and expert systems, have been widely applied for image classification. Generally speaking, image classification approaches can be grouped as supervised and unsupervised, or parametric and nonparametric, or hard and soft classification, which is detailed in [10, 11, 28, 29].

3.1. Combination of multiple classifiers

Different classifiers, such as parametric classifiers and non-parametric classifiers have their own strengths and limitations [10, 11]. For example, when sufficient training samples are available and the feature of land covers in a dataset is normally distributed, a maximum likelihood classifier (MLC) may yield an accurate classification result. In contrast, when image data are anomalously distributed, neural network and decision tree classifiers may demonstrate a better classification result [8, 30, 31]. Previous research has indicated that the integration of two or more classifiers provides improved classification accuracy compared to the use of a single classifier. A critical step is to develop

suitable rules to combine the classification results from different classifiers. Some previous research has explored different techniques, such as a production rule, a sumrule, stacked regression methods, majority voting, and thresholds, to combine multiple classification results.

4. Use of Multiple Features of Remote-Sensing Data

As discussed previously, remote-sensing data have many unique spatial, spectral, radiometric, temporal, and polarization characteristics. Making full use of these characteristics is an effective way to improve classification accuracy. In general, the feature of spectral response is the most important information used for land-cover classification. As high spatial resolution data become readily available, textural and contextual information become significant in image classification.

Spatial resolution determines the level of spatial detail that can be observed on the Earth's surface. As fine spatial resolution data, such as IKONOS and QuickBird, become more easily available, they are increasingly employed for different applications. A major advantage of these fine spatial resolution images is that such data greatly reduce the mixed-pixel problem, providing a greater potential to extract much more detailed information on land-cover structures than medium or coarse spatial resolution data. However, some new problems associated with fine spatial resolution image data emerge, notably the shadows caused by topography, tall buildings, or trees, and the high spectral variation within the same land-cover class. These disadvantages may lower classification accuracy if classifiers cannot effectively handle them. Increased spectral variation is common with the high degree of spectral heterogeneity in complex landscapes. The huge amount of data storage and severe shadow problems in fine spatial resolution images lead to challenges in the selection of suitable image-processing approaches and classification algorithms. Last, but not least, high spatial resolution

imagery is much more expensive and requires much more time to implement data analysis than medium spatial resolution images. In order to make full use of the rich spatial information inherent in fine spatial resolution data, it is necessary to minimize the negative impact of high intraspectral variation. Spatial information may be used in different ways, such as in contextual-based or object-oriented classification approaches, or classifications with textures. The combination of spectral and spatial classification is especially valuable for fine land-cover classification systems in the areas with complex landscapes. As contextual-based and object-oriented classification approaches have been discussed previously, the following only focuses on the use of textures in image classification. In addition, many texture measures have been used for image classifications [32, 33].

5. Conclusions

Image classification has made great progress over the past decades in the following three areas: (1) development and use of advanced classification algorithms, such as subpixel, per-field, and knowledge-based classification algorithms; (2) use of multiple remote-sensing features, including spectral, spatial, multitemporal, and multisensor information; and (3) incorporation of ancillary data into classification procedures, including such data as topography, soil, road, and census data.

Accuracy assessment is an integral part in an image classification procedure. Accuracy assessment based on error matrix is the most commonly employed approach for evaluating per-pixel classification, while fuzzy approaches are gaining attention for assessing fuzzy classification results. Uncertainty and error propagation in the image-processing chain is an important factor influencing classification accuracy. Identifying the weakest links in the chain and then reducing the uncertainties are critical for improvement of classification accuracy. The study of uncertainty will be an important topic in the future research of image classification.

Spectral features are the most important information for image classification. As spatial resolution increases, texture or context information becomes another important attribute to be considered. Classification approaches may vary with different types of remote-sensing data. For example, with high spatial resolution data, the severe impact of the shadow problem resulting from topography and vegetation stand structures and the wide spectral variation within the land-cover classes may outweigh the advantages from high spatial resolution if a per-pixel, spectral-based classification is used for these image classifications. Under this circumstance, a combination of spectral and texture information can reduce this problem and per-field or object-oriented classification algorithms outperform per-pixel classifiers.

Although spatial information is remarkably useful for fine spatial resolution data, how to effectively derive and use it in image classification remains a research topic. Texture, shape, and context information are currently most frequently used. However, even with the most widely used texture information, there is still much uncertainty in the determination of texture measures, image channel, window size, and other parameters. More research is necessary to develop a guideline for selecting textures suitable for different biophysical environments.

The success of an image classification depends on many factors. The availability of high-quality remotely sensed imagery and ancillary data, the design of a proper classification procedure, and the analyst's skills and experiences are the most important ones. For a particular study, it is often difficult to identify the best classifier due to the lack of a guideline for selection and the availability of suitable classification algorithms to hand. Comparative studies of different classifiers are thus frequently conducted. Moreover, the combination of different classification approaches has shown to be helpful for improvement of classification accuracy. It is necessary for future research to develop guidelines on the applicability and capability of major classification algorithms.

Acknowledgements

This work is supported by the Graduates' Research Innovation Program of Higher Education of Jiangsu Province (Grant No. CXZZ13-0239).

References

- [1] P. Gong and P. J. Howarth, Frequency-based contextual classification and gray-level vector reduction for land-use identification, *Photogrammetric Engineering and Remote Sensing* 58 (1992), 423-437.
- [2] C. Kontoes, G. G. Wilkinson, A. Burrill, S. Goffredo and J. Megier, An experimental system for the integration of GIS data in knowledge-based image analysis for remote sensing of agriculture, *International Journal of Geographical Information Systems* 7 (1993), 247-262.
- [3] G. M. Foody, Approaches for the production and evaluation of fuzzy land cover classification from remotely-sensed data, *International Journal of Remote Sensing* 17 (1996), 1317-1340.
- [4] M. A. J. San and G. S. Biging, Comparison of single-stage and multi-stage classification approaches for cover type mapping with TM and SPOT data, *Remote Sensing of Environment* 59 (1997), 92-104.
- [5] P. Aplin, P. M. Atkinson and P. J. Curran, Per-field classification of land use using the forthcoming very fine spatial resolution satellite sensors: Problems and potential solutions, In P. M. Atkinson and N. J. Tate (Eds.), *Advances in Remote Sensing and GIS Analysis* (1999), 219-239 (New York: John Wiley and Sons).
- [6] J. Stuckens, P. R. Coppin and M. E. Bauer, Integrating contextual information with per-pixel classification for improved land cover classification, *Remote Sensing of Environment* 71 (2000), 282-296.
- [7] S. E. Franklin, D. R. Peddle, J. A. Dechka and G. B. Stenhouse, Evidential reasoning with Landsat TM, DEM and GIS data for landcover classification in support of grizzly bear habitat mapping, *International Journal of Remote Sensing* 23 (2002), 4633-4652.
- [8] M. Pal and P. M. Mather, An assessment of the effectiveness of decision tree methods for land cover classification, *Remote Sensing of Environment* 86 (2003), 554-565.
- [9] F. J. Gallego, Remote sensing and land cover area estimation, *International Journal of Remote Sensing* 25 (2004), 3019-3047.
- [10] B. Tso and P. M. Mather, *Classification Methods for Remotely Sensed Data*, Taylor and Francis Inc., New York, 2001.
- [11] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*, John Wiley and Sons, Hoboken, NJ, 2003.

- [12] L. Hubert-Moy, A. Cotonne, L. Le Du, A. Chardin and P. Perez, A comparison of parametric classification procedures of remotely sensed data applied on different landscape units, *Remote Sensing of Environment* 75 (2001), 174-187.
- [13] D. Chen and D. A. Stow, The effect of training strategies on supervised classification at different spatial resolution, *Photogrammetric Engineering and Remote Sensing* 68 (2002), 1155-1162.
- [14] P. M. Mather, *Computer Processing of Remotely-Sensed Images: An Introduction*, 3rd Edition, John Wiley & Sons, Chichester, 2004.
- [15] G. F. Hughes, On the mean accuracy of statistical pattern recognizers, *IEEE Transactions on Information Theory* 14 (1968), 55-63.
- [16] K. P. Price, X. Guo and J. M. Stiles, Optimal Landsat TM band combinations and vegetation indices for discrimination of six grassland types in eastern Kansas, *International Journal of Remote Sensing* 23 (2002), 5031-5042.
- [17] S. W. Myint, A robust texture analysis and classification approach for urban land-use and land-cover feature discrimination, *Geocarto International* 16 (2001), 29-40.
- [18] G. S. Okin, D. A. Roberts, B. Murray and W. J. Okin, Practical limits on hyperspectral vegetation discrimination in arid and semiarid environments, *Remote Sensing of Environment* 77 (2001), 212-225.
- [19] T. Rashed, J. R. Weeks, M. S. Gadalla and A. G. Hill, Revealing the anatomy of cities through spectral mixture analysis of multispectral satellite imagery: A case study of the Greater Cairo region, Egypt, *Geocarto International* 16 (2001), 5-15.
- [20] G. P. Asner and K. B. Heidebrecht, Spectral unmixing of vegetation, soil and dry carbon cover in arid regions: Comparing multispectral and hyperspectral observations, *International Journal of Remote Sensing* 23 (2002), 3939-3958.
- [21] D. B. Lobell, G. P. Asner, B. E. Law and R. N. Treuhaft, View angle effects on canopy reflectance and spectral mixture analysis of coniferous forests using AVIRIS, *International Journal of Remote Sensing* 23 (2002), 2247-2262.
- [22] R. A. Neville, J. Levesque, K. Staene, C. Nadeau, P. Hauff and G. A. Borstad, Spectral unmixing of hyperspectral imagery for mineral exploration: Comparison of results from SFSI and AVIRIS, *Canadian Journal of Remote Sensing* 29 (2003), 99-110.
- [23] R. V. Platt and A. F. H. Goetz, A comparison of AVIRIS and Landsat for land use classification at the urban fringe, *Photogrammetric Engineering and Remote Sensing* 70 (2004), 813-819.
- [24] J. Cihlar, Q. Xiao, J. Chen, J. Beaubien, K. Fung and R. Latifovic, Classification by progressive generalization: A new automated methodology for remote sensing multispectral data, *International Journal of Remote Sensing* 19 (1998), 2685-2704.
- [25] G. M. Foody, Status of land cover classification accuracy assessment, *Remote Sensing of Environment* 80 (2002), 185-201.

- [26] D. K. McIver and M. A. Friedl, Estimating pixel-scale land cover classification confidence using nonparametric machine learning methods, *IEEE Transactions on Geoscience and Remote Sensing* 39 (2001), 1959-1968.
- [27] W. Liu, S. Gopal and C. E. Woodcock, Uncertainty and confidence in land cover classification using a hybrid classifier approach, *Photogrammetric Engineering and Remote Sensing* 70 (2004), 963-971.
- [28] J. Cihlar, Land cover mapping of large areas from satellites: Status and research priorities, *International Journal of Remote Sensing* 21 (2000), 1093-1114.
- [29] S. E. Franklin and M. A. Wulder, Remote sensing methods in medium spatial resolution satellite data land cover classification of large areas, *Progress in Physical Geography* 26 (2002), 173-205.
- [30] D. Lu, P. Mausel, M. Batistella and E. Moran, Comparison of land-cover classification methods in the Brazilian Amazon Basin, *Photogrammetric Engineering and Remote Sensing* 70 (2004), 723-731.
- [31] D. Lu and Q. Weng, A survey of image classification methods and techniques for improving classification performance, *International Journal of Remote Sensing* 28(5) (2007), 823-870.
- [32] D. K. Gordon and W. R. Phillipson, A texture enhancement procedure for separating orchard from forest in Thematic Mapper imagery, *International Journal of Remote Sensing* 8 (1986), 301-304.
- [33] S. E. Franklin and D. R. Peddle, Spectral texture for improved class discrimination in complex terrain, *International Journal of Remote Sensing* 10 (1989), 1437-1443.

