Textural-Contextual Labeling and Metadata Generation for Remote Sensing Applications

IN-43
045652

Richard K. Kiang*
NASA Goddard Space Flight Center
Greenbelt, MD 20771

ABSTRACT

Despite the extensive research and the advent of several new information technologies in the last three decades, machine labeling of ground categories using remotely sensed data has not become a routine process. Considerable amount of human intervention is needed to achieve a level of acceptable labeling accuracy. A number of fundamental reasons may explain why machine labeling has not become automatic. In addition, there may be shortcomings in the methodology for labeling ground categories. The spatial information of a pixel, whether textural or contextual, relates a pixel to its surroundings. This information should be utilized to improve the performance of machine labeling of ground categories. Landsat-4 Thematic Mapper (TM) data taken in July 1982 over an area in the vicinity of Washington, D.C. are used in this study. On-line texture extraction by neural networks may not be the most efficient way to incorporate textural information into the labeling process. Texture features are pre-computed from cooccurrence matrices and then combined with a pixel's spectral and contextual information as the input to a neural network. The improvement in labeling accuracy with spatial information included is significant. The prospect of automatic generation of metadata consisting of ground categories, textural and contextual information is discussed.

Keywords: neural network, remote sensing, textural, contextual, metadata

1. INTRODUCTION

The first earth resource satellite was launched in 1972 with a 4-channel Multispectral Scanner System on board. With this first instrument, the vast potential of monitoring the condition of the earth's environment from space was discovered. Since then, the remote sensing technology has grown significantly. The first Earth Observing System platform with five instruments will be launched in 1999. As smaller missions with less expensive sensors may become the trend for spaceborne remote sensing, measurements from a fleet of advanced sensors will be used collectively to derive the physical parameters for the land, the oceans, and the atmosphere, for monitoring climate changes and understanding the underlying mechanisms. To process, analyze, archive, and distribute this large amount of data in a timely manner not only presents a tremendous challenge to the data systems for these sensors, but also demands innovative analytical methods and advanced computing and data communication technologies.

Despite the extensive research and the advent of several new information technologies in the last three decades, machine labeling of ground categories using remotely sensed data has not become a routine process. Considerable amount of human intervention is needed to achieve a level of acceptable labeling accuracy. There are a number of fundamental reasons why machine labeling has not become automatic: (1) The spectral response of a ground category as viewed by the sensor is not constant. It depends on the viewing geometry, conditions of the atmosphere and the sensor, and geophysical properties of the target pixel and the surrounding pixels. Since each ground category has variable and sometimes unpredictable spectral response, the inversion process to infer ground category from spectral response is understandably difficult. (2) Ground truth is needed in training and testing any labeling algorithms. Since ground truth information must be collected manually and a ground truth map must be constructed with human efforts, it is tedious and expensive to produce such maps, and it is not uncommon to find labeling errors in a ground truth. Shortage of good, reliable ground truth hampers the development of machine labeling algorithms. (3) The sensor's spectral or spatial characteristics may not be suitable for identifying certain ground categories. The development of hyperspectral instruments with one or two magnitudes more channels than the early Landsat would indeed provide the appropriate spectral channels to detect certain geophysical events. However, machine labeling using hyperspectral data has other complications.[11]

* Correspondence: E-mail: richard.kiang@gsfc.nasa.gov

In addition, there may be shortcomings in the methodology for labeling ground categories. Photointerpretation by human analysts in general has superior performance than computers. Therefore machine labeling algorithms may benefit from emulating how photointerpretation is performed. Photoanalysts use at least as much spatial information as the spectral information for segmentation, detection, identification and interpretation. The majority of the research in machine labeling, however, does not take spatial information into consideration. In many algorithms only one pixel is considered at a time; pixel location does not matter because pixel coordinates never enter the computation. The spatial information of a pixel, whether textural or contextual, relates the relationship a pixel bears with its neighbors. The information must be utilized to improve the performance of machine labeling of ground categories.

Metadata is the data accompanying the dataset to provide the essential information of the dataset. It may describe how the dataset is generated or provide information of the important features or events in the dataset. As the data volume available to the users grows exponentially in the future, query on metadata is an essential means of screening datasets without actually examining the content of the dataset. Metadata may indicate the availability of certain features (e.g. deciduous forest, water bodies) and presence of certain events (e.g. fires, storms). In general, automated techniques should be used to extract the needed information from the dataset and used as the metadata. Ground categories and spatial information are useful information that can be part of the metadata.

In the following, Section 2 reviews certain textural and contextual features in image processing, and explains the features used in the analysis; Section 3 describes the data and the ground truth used in the study; Section 4 presents the training process and the classification results; Section 5 suggests how metadata can be extracted in this process; and Section 6 discusses the results and suggests future directions.

2. TEXTURAL AND CONTEXTUAL FEATURES

There is no consensus on the precise definition of texture, in spite of the extensive research conducted in the last three decades. Conceptually, texture can be considered as an macroscopic attribute generated by repetitive primitives according to a placement rule.

Various methods have been used for textural analysis, modeling and synthesis in the past. For example, Haralick et al.[9] used features extracted from gray-tone spatial-dependence matrices to classify a photomicrograph, an aerial photograph, and a satellite image. Weszka et al.[17] used features from Fourier power spectrum, second-order gray level statistics, and first-order statistics of gray level differences to classify Landsat imagery samples. Tamura et al.[15] proposed to use six features — coarseness, contrast, directionality, line-likeness, regularity, and roughness — obtained from second-order statistics for texture analysis. Haralick[8] reviewed the approaches and models investigators have used for texture analysis. He concluded that the statistical approaches work better for microtextures; and for macrotextures, histograms of primitive properties and cooccurrence of primitive properties may be used. Conners et al.[5] concluded that the spatial gray level dependence method (also called concurrence matrix) performed better than the gray level run length method, the gray level difference method, or measures derived from Fourier spectrum using generated textures. Pentland[13] used fractal functions to represent natural shapes such as mountains, trees, and clouds. Texture was modeled by an isotropic fractal Brownian function with constant fractal dimension. Chellappa et al.[4] used Markov random field models to generate textures. Derin et al.[7] used Gibbs distribution for modeling and segmentation of noisy textured images, and presented a dynamic programming segmentation algorithm. Bischof et al.[3] appended TM Band 5 measurements from 5x5 or 7x7 neighborhoods to each pixel and relied on a neural network to extract textural information. A two-layer network was also used in smoothing postprocessing for each 5x5 neighborhood. Rao et al.[14] discovered that three high-level features, namely repetition, orientation and complexity, are important to attentive texture perception. Augusteijn et al.[2] used five categories of texture measures to classify segments of TM data. These five categories are: cooccurrence matrices, gray-level differences, texture-tone analysis, feature derived from the Fourier spectrum, and Gabor filters. Some Fourier features and some Gabor filters were considered to be good choices especially when only one band was used for classification. Ojala et al.[12] evaluates the performance of such texture measures as gray-level difference methods, Laws' texture measures, center-symmetric covariance measures, and local binary patterns. Zhu et al.[20] used features from wavelet transform to classify relief images from aerial photographs digitized at various resolutions. The advantages of multiresolution analysis over other traditional approaches for this type of applications were discussed. For spaceborne remote sensing, this implies that texture features for sensors of different spatial resolutions can be used across sensor platforms. From these previous investigations, it appears that there is no single method that is suitable for all applications. The most appropriate approach is driven by the application itself.

In this study, texture measures computed from the cooccurrence matrix are used, along with other textural-contextual indicators such as the mean values and the standard deviations of the 3x3 and the 5x5 neighborhoods. It could be argued that the texture measures could be extracted automatically from a neural network and hence they do not need to be precomputed. However, testings indicate that this may not be the optimum approach as many texture features may be used. Therefore the textural measures are computed off-line.

A cooccurrence matrix $\{P_{ij}(d,\theta)\}$ consists of relative frequencies P_{ij} with which two neighboring pixels separated by distance d in direction θ occur on the image, one with level i and the other with level j. Energy, uniformity, or homogeneity is defined as ΣP_{ij}^2. Entropy is defined as $-\Sigma P_{ij}\log P_{ij}$. And contrast is defined as $\Sigma(i-j)^2 P_{ij}^2$. For this study, d=1, and θ = 0°, 45°, 90°, 135°.

In a contextual classification, one can use information derived from the context of the data or ancillary information associated with the data to aid the classification. For example, in machine reading postal addresses on envelopes, the street name and the number, the city name, the state name and the zip code must all be consistent. This constraint can greatly improve the recognition rate. For remote sensing applications, contextual classification is less developed and used infrequently comparing with textural classification. A common contextual rule in classifying natural scenes is that a pixel is more likely to have the same ground category as the majority of its neighboring pixels have. This is based on the fact that the scales of ground categories in natural scenes are much larger than a pixel's spatial resolution. Post-processing majority filters can be constructed to rectify pixels that appear to be misclassified. A different approach proposed by Wharton[18] to improve classification accuracy is a two-pass contextual classification based on ground category distribution in a local area. In reality, any reasonable rules, such as "nearly all the agricultural fields in this county are rectangular" may be used to improve classification accuracy. In this paper, contextual information is provided by the local statistics (mean and standard deviation) in the 3x3 and the 5x5 neighborhoods. Pixel counts for the individual pixels in the 3x3 or the 5x5 neighborhoods in one or more bands may be concatenated to the center pixel also for the neural network to extract textural and contextual information. However, as will be discussed in Section 4, the exact nature of textural and contextual information extracted by a neural network may not be easily understood. In summary, the spatial features that could be included in training and classification are four sets of the parameters computed from the cooccurrence matrices — energy, entropy and contrast for the four orientations, and two sets of mean and standard deviation for the 3x3 and the 5x5 neighborhoods. In addition, pixel counts from any of the TM bands in the 3x3 or the 5x5 neighborhood may also be included.

3. CHARACTERISTICS OF THE DATA USED IN THE STUDY

Landsat-4 Thematic Mapper (TM) data taken in July 1982 over an area in the vicinity of Washington, D.C. were used in this study. Only the first four TM bands were available, as the instruments for the three remaining IR bands had not stabilized. The ground truth consists of 17 categories, and were obtained through photointerpretation of color infrared aerial photographs and subsequent field visits.[19]

In general, ground truth contains information categories instead of spectral categories. Since the neural networks in this study perform classifications based on spectral data, whether the information categories correspond to distinct spectral categories should be examined to estimate the intrinsic discriminability among the categories. To achieve this objective, the spectral signatures for all categories are computed. The signatures consist of means vectors and covariance matrices. A number of measures, such as divergence and Mahalanobis distance, could be used to estimate the separability among multi-dimensional clusters. In this study, we compute the ratio of between-class variance to within-class variance along the Fisher optimal discriminant vector.[6] From these ratios, it is concluded that some information categories are heavily overlapped with others, and that the 17 information categories should be combined into six categories, following the land use and land cover classification system of Anderson et al.[1] These six categories are: (1) urban or built-up land, (2) agricultural land, (3) rangeland, (4) forest land, (5) water, and (7) bare soil /cleared land. Notice that there is no Category 6 — wetland — in this data. In Anderson's system, Category 7 is barren land, such as salt flats, beaches, bare rock, etc. Since bare soil/cleared land (Category 17 in the ground truth data) does not exactly fit the definition, the original description in the ground truth is used instead.

It has been noticed by Telfer et al.,[16] however, that the ground truth data had certain errors in it. Since a ground truth is normally constructed with manual processes, it is not uncommon that the ground truth is inherited with some human errors. It should be kept in mind also that because of the pixel's finite spatial resolution a ground truth is never perfect even when there is no human error.

4. TRAINING AND CLASSIFICATION

A 3-layer, feed-forward neural network is used in this study. The input layer has four or more units. The four units corresponding to the four TM bands. Additional units are used for the textual or contextual input features. The hidden layer has ten units, and the output layer has six units, corresponding to the six ground categories. The area has 21,952 pixels with defined ground truth. Excluding the edge pixels because they do not have a complete 5x5 neighborhood, there are 20,441 usable pixels, among which half of the pixels are used for training, and the other half for testing. Using alternate pixels for training and testing is not a normal classification strategy. The purpose of using it here is to explore the optimum classification accuracy for testing samples, especially there could be complications with the increased dimensionality (see Section 6). Because the two sample pools have virtually indistinguishable distributions, the testing accuracy can sometimes be slightly higher than the training accuracy.

Small learning rate and momentum factor are used in the training to minimize fluctuation. Convergence is usually reached within several thousand iterations. The classification accuracies (percent correctly identified) for the training and the testing samples are tabulated as follows.

Category	Training	Testing
1	78.5	80.4
2	63.5	65.5
3	14.1	13.8
4	84.7	84.4
5	80.0	69.2
7	62.2	60.4
Overall	71.5	71.6

If samples classified with lower activation values are rejected to enhance the confidence level of the classification, then the classification accuracies can be improved. Previously it was shown that by rejecting 9.7% of the samples with lower activation the overall classification can be improved to 75.4%. On the average, for each percent increase of accuracy, approximately 2.3% of the samples with lower activation values are rejected.[10]

Classification accuracies with spatial features included as shown in the following table. *mean3* and *sd3* are the mean and the standard deviation for the 3x3 neighborhood. *mean5* and *sd5* are those for the 5x5 neighborhood. It is apparent that TM3 contains more spatial information for discrimination than TM1.

	Overall Classification Accuracy (%)	
	Training Samples	Testing Samples
TM1–4 pixel's spectral data only	71.5	71.6
Plus TM1–4 mean3	71.2	69.9
Plus TM1–4 sd3	63.2	62.7
Plus TM1–4 mean3 and sd3	77.0	76.6
Plus TM1–4 mean3, sd3, mean5 and sd5	68.8	68.3
Plus TM1–4 energy	68.4	67.5
Plus TM1–4 entropy	74.0	73.0
Plus TM1–4 contrast	75.8	75.4
Plus TM1 energy, entropy, contrast	73.9	74.0
Plus TM1 3x3 neighborhood	68.6	68.1
Plus TM1 5x5 neighborhood	62.5	63.0
Plus TM1 energy, entropy, contrast, and 5x5 neighborhood	69.4	69.0
Plus TM3 energy, entropy, contrast	73.6	73.5
Plus TM3 3x3 neighborhood	72.9	72.1
Plus TM3 5x5 neighborhood	75.5	75.6
Plus TM3 energy, entropy, contrast, and 5x5 neighborhood	76.4	76.4

Plus TM1 and TM3 energy, entropy, contrast, and 5x5 neighborhood	76.9	76.5

5. METADATA GENERATION

Metadata is loosely defined as "the data about data", or "the additional information that is necessary for data to be useful." Most elements in computer-based applications, such as datasets, databases, data warehouses, software, and computer systems, have metadata as components. It provides the information concerning data's structure, context and meaning. Metadata management issues, such as enterprise-wide maintenance and integration, and interoperability among internal and external elements, are important aspects in modern computer systems.

For datasets in remote sensing applications, metadata may describe how a dataset is generated or provide information of the important features or events in the dataset. Metadata may indicate the availability of certain features (e.g. deciduous forest, water bodies) and presence of certain events (e.g. fires, storms). Ground categories and spatial characteristics are useful information that can be included as metadata also.

As the amount of remotely sensed data will increase exponentially in the future, a challenge facing the users and the data providers alike is to determine which datasets a user will really need for a specific application. It will be impractical for a user to request a large number of datasets, download them to the user's computing platform, process the datasets, then determine which datasets should be used. From a user's point of view, it is desirable to have a screening capability to reduce the number of datasets that have to be examined. Browsing a low-resolution image can be a part of the screening process. But more detailed information, such as the presence of certain ground categories and textures, or the approximate ratios of the categories, may also be useful indices for searching. A solution is to extract features and events in a dataset automatically as a normal step along the generation of the standard Level-1 datasets. The features and events can then be included in the metadata. Queries including more specific information would greatly limit the amount of datasets for examination.

6. DISCUSSIONS

With spatial features included, the labeling accuracies could improve by up to approximately 6%. This is a significant increase considering the accuracy is approximately 71% without spatial features. In the study, a fixed network architecture with fixed number of hidden units is used in all trainings in order to reveal the impact of inclusion of spatial features. Since in some cases the number of input features increases by several folds after the spatial features are included, conceivably the training and the testing accuracies may increase further if more hidden units are used.

It is not efficient to rely on the network to compute a whole set of statistical texture measures and then perform labeling process. This is the rationale to compute energy, entropy, contrast, and neighborhood statistics separately. By appending the neighboring pixel to the center pixel spatial features are nevertheless extracted. It would be interesting to explore the nature of the textural or contextual features so extracted.

As spatial features are included, the dimensionality of the labeling problem could increase rapidly. Since the number of training samples remains unchanged, the feature space with the enhanced dimensionality will be largely empty. Since there are very large gaps among pixels in the feature space, the decision boundary being determined by the neural network may have very large leeway. This implies that during the training process RMS error may decrease without realizing any increase in labeling accuracy. The large leeway may lead to poor generalization and performance if the training and testing samples do not belong to the same population. Increasing the size of the training sample is the only way to reduce the gaps in the feature space. However, as pointed out in Section 1, this is an expensive proposition because ground truth is difficult to obtain.

7. REFERENCES

1. J. R. Anderson et al., "A land use and land cover classification system for use with remote sensor data," Geological Survey Professional Paper No. 964, 1976. Also in *Manual of Remote Sensing*, 2nd Edition, Table 30-7, American Society of Photogrammetry, 1983.

2. M. F. Augusteijn, L. E. Clemens, and K. A. Shaw, "Performance evaluation of texture measures for ground cover identification in satellite images by means of a neural network classifier," *IEEE Trans. on Geoscience and Remote Sensing,* Vol. 33, No. 3, pp.616–626, May 1995.
3. H. Bischof, W. Schneider, and A. J. Pinz, "Multispectral classification of Landsat images using neural networks," *IEEE Trans. Geoscience Remote Sensing,* Vol. 30, No. 3, pp. 482-490, May 1992.
4. R. Chellappa, S. Chatterjee, "Classification of textures using Gaussian Markov random fields," *Trans. ASSP*, Vol. 33, No. 4, pp. 959-963, 1985.
5. R. W. Conners and C. A. Harlow, "A theoretical comparison of texture algorithms," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* Vol. PAMI-2, No. 3, pp.204–222, May 1980.
6. P. W. Cooper, "Hyperplanes, hyperspheres, and hyper-quadrics as decision boundaries," in *Computer and Information Sciences,* eds. J. T. Tou and R. H. Wilcox, Spartan Books, Washington, D.C, 1964.
7. H. Derin and H. Elliott, "Modeling and segmentation of noisy and textured images using Gibbs random fields, " *IEEE Trans. on Pattern Analysis and Machine Intelligence,* Vol. 9, pp. 39-55, 1987.
8. R. M. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE,* Vol. 67, No. 5, pp. 786–804, May 1979.
9. R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. on System, Man, and Cybernetics,* Vol. SMC-3, No. 6, pp. 610–621, November 1973.
10. R. K. Kiang, "Classification of remotely sensed data using OCR-inspired neural network techniques," in the Proceedings of International Geoscience and Remote Sensing Symposium, Houston, TX, pp. 1081–1083, May 1992.
11. L. O. Jimenez and D. A. Landgrebe, "Supervised classification in high dimensional space: geometrical, statistical, and asymptotical properties of multivariate data," *IEEE Trans. on System, Man, and Cybernetics,* Vol. 28, Part C, No. 1, pp. 39–54, Feb. 1998.
12. T. Ojala, M. Pietikainen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," Pattern Recognition, Vol. 29, No. 1, pp. 51–59, 1996.
13. A. P. Pentland, "Fractal-Based Description of Natural Scenes," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* Vol. PAMI-6, No. 6, November 1984.
14. R. A. Rao and G. L. Lohse, "Identifying high level features of texture perception," *CVGIP: Graphical Models and Image Processing,* Vol. 55, pp. 218–233, 1993.
15. H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception," *IEEE Trans. on Systems, Man, and Cybernetics,* Vol. SMC-8, pp. 460-473, 1978.
16. B. A. Telfer, H. H. Szu, and R. K. Kiang, "Classifying multispectral data by neural networks," *Telematics and Informatics,* Vol. 10, No. 3, pp. 209–222, 1993.
17. J. S. Weszka, C. R. Dyer, and A. Rosenfeld, "A comparative study of texture measures for terrain classification," *IEEE Trans. Syst., Man, Cybern.,* Vol. SMC-6, pp. 269–285, 1976.
18. S. W. Wharton, "A contextual classification method for recognizing land use patterns in high resolution remotely sensed data," *Pattern Recognition,* Vol. 15, No. 4, pp. 317–324, 1982.
19. D. L. Williams et al., "A statistical evaluation of the advantages of Landsat Thematic Mapper data in comparison to MSS data," *IEEE Trans. on Geoscience and Remote Sensing,* GE-22, pp.294–302, 1984.
20. C. Zhu and X. Yang, "Study of remote sensing image texture analysis and classification using wavelet," *Int. J. Remote Sensing,* Vol. 19, No. 16, pp. 3197–3203, 1998.