Journal of the Korean Society of Remote Sensing, Vol. 8, No. 2, 1992

A Discussion on the Approaches for Interfacing Remote Sensing and Geographic Information Systems

Song Hak Choung* and Kap-Duk Kim
*Dept. of Forest Management, Forestry Research Institute
Dongdaemun, Seoul 130-012, Korea
(Received November 10, 1992 ; Accepted November 20, 1992)

遠隔探査와 地理情報시스템間의 接木方法에 관한 考察

鄭聖鶴*·金甲德
*임업연구원 산림경영부 자원관리과
(1992년 11월 10일 받음 ; 1992년 11월 20일 수리)

要 約

원격탐사와 지리정보시스템은 많은 분야에서 접목되어 활용되고 있다. 이러한 두 공간자료처리시스템간에 자료의 이동방법에 관하여 두 가지 기법을 고찰하였다. 원격탐사자료를 이용하여 자연자원을 정확하게 구분하는 데에는 어려움이 따른다. 그 정확도를 높이기 위해서는 보조자료, 즉 디지타이즈드된 지도 및 지형(고도)자료 등을 원격탐사자료와 결합하여 이용한다. 이러한 자료를 이용하는 데에는 (1)구분 전 층화와 (2)구분 후 정리 등의 두 가지 기법이 많이 쓰인다. 이 두 기법은 유용한 반면, 결정규칙에 의존함으로써 다소 전문성이 결여된다.

Abstract

Interconnecting remote sensing systems to geographic information systems is valuable in many different applications. Two common techniques for moving data between these two related kinds of spatial data-processing systems were discussed. Digital classification of remote sensing data for use in natural resource inventory has produced mixed results. In attempts to improve classification, accuracy ancillary data, such as digitized maps and terrain(elevation) data, have been combined with remotely sensed data in various ways. These data have been used commonly in (1) preclassification scene stratification and (2) postclassification class sorting. These two approaches are found to be efficient, but lacking in sophistication due to their reliance on deterministic decision rules.

Introduction

Remote sensing often requires other kinds of ancillary data to achieve both its greatest value and the highest levels of accuracy as a data and information production technology. Geographic information systems can provide this capability. They permit the integration of datasets acquired from library, laboratory, and fieldwork with remotely sensed data(Star and Estes, 1990). On the other hand, applications of GIS are heavily dependent on both the timeliness or currency of the data they contain, as well as the geographic coverage of the database. For a variety of applications, remote sensing, while only one source of potential input to a GIS, can be valuable. Remote sensing can provide timely data at scales appropriate to a variety of applications. As such, many researchers feel that the use of geographic information systems and remote sensing can lead to important advances in research and operational applications. Merging these two technologies can result in a tremendous increase in information for many kinds of users.

In a complementary fashion, image processing and interpretation of remotely sensed data must employ ancillary or collateral data(such as elevation data and existing land-use/land-cover data) to achieve high levels of thematic classification accuracy, which naturally brings us to work in the GIS environment(Star and Estes, 1990). That is, improvement in digital classification of remote sensing data can be achieved through the incorporation of ancillary data. These data may be choroplethic maps of various land attributes or digital terrain(elevation) data(Hutchinson, 1982). The incorporation of ancillary data in the classification process can be approached in several ways. The general characteristics of these approaches are described.

Approaches to Use of Ancillary Data

1. Stratification

Use of ancillary data prior to classification involves a division of the study scene into smaller areas or strata based on some criterion or rules, so that each stratum may be processed independently. Statistically, the purpose of stratification is to increase the homogeneity of the data sets to be classified(Hutchinson, 1982). Because of its simplicity, stratification is a widely used technique. From a practical standpoint, stratification is employed for classification improvement either to divide a large study area into smaller homogeneous units, or to separate different things which are spectrally similar.

There are two advantages to dividing a large study area into smaller subareas. First is the simple convenience of dealing with smaller data sets at each stage of analysis. This, in fact, may be an overriding practical consideration in especially large studies(Bryant et al., 1979). The second advantage is a reduction of variant within strata. This is the statistical basis for stratification. The spectral characteristics of any set of objects, such as specific soil or vegetation types, are likely to vary over distance. As variance increases, the likelihood of confusion between spectrally similar objects also increases. Criteria selected for stratification should be significant in describing the variation of the objects of interest within the study area. For example, a regional study of soils might be stratified by rock type, or a vegetation study might be stratified by elevation.

A more specific and pragmatic application of stratification is its use in separating different objects, obviously different things, such as older residential areas and rural woodlands, may be spectrally identical. To avoid confusion, urban and non-urban areas may be separated by manual photo-interpretation or by using a general land-use map. Training and classification then can proceed independently on each stratum and finally the two may be merged in a final product. Confusion is thus avoided and accuracy improved(Gaydos and Newland, 1978).

Stratification is a conceptully simple tool and, carefully used, can be effective in improving classification accuracy. However, it is not sensitive to subtle distinctions. Differences between strata are absolute and the lines between them are abrupt ; there are no gradations or fuzzy boundaries between mapped classes(Hutchinson, 1982). Thus, considerable care should be taken when (1) deciding to stratify, or not and (2) selecting stratification criteria.

Imprudent selection of stratification criteria can have far-reaching implications in classification. Differences in training set selection for individual strata and/or the vagaries of clustering algorithms, if used, may produce markedly different spectral classes on either side of strata boundaries. Merging strata for a final product which class boundary offsets or missing classes is difficult, at best.

2. Postclassification Sorting

The use of ancillary data after multispectral classifcation is based on the observation that a single class of objects seldom can be represented by a single spectral class. To accommodate this, a large number of spectral classes commonly is created. Spectral classes may then be merged into groups which represent object classes. The problem is that one spectral class may often represent subsets of more than one object class. In postclassification sorting, these problem spectral classes are treated as separate special cases. Based on a sorting rule, individual pixels of the problem spectral class are assigned to the appropriate object class using ancillary data. The approach and techniques used in postclassification sorting are derived from methods for overlay analysis found in grid-based geographic information systems.

3. Advantages and Disadvantages

(1) Stratification

The use of ancillary data for scene stratification has been widely used in many different types of applications. Stratification is statistically sound, easily implemented, effective, and inexpensive in computer time. However, it is deterministic and thus cannot accommodate gradations between strata. In addition, because it is performed before classification, incorrect stratification criteria can invalidate the entire classification.

(2) Postclassification Sorting

This is a rather new application of a technique that has been used in geographic information systems. It, like stratification, is conceptually simple and easily implemented, but it is also deterministic. However, it does offer some advantages: it deals only with problem classes rather than all classes and, unlike stratification, errors made in the selection of sorting rules are easily corrected and do not require that the classification be redone.

Discussion and Conclusion

Each of the techniques for classification improvement has advantages that will recommend its use in particular situations. Thus, it is not useful or proper to offer a judgement as to what the "best" technique might be. However, it is appropriate to make some general observations. Because of their simplicity, stratification and postclassification sorting will likely continue to be used in spite of their limitations. Both techniques are most effective when the confused objects are relatively discrete in their distribution, as is the case in many urban applications.

The simple addition of logical channels is difficult to recommend because results cannot be consistently predicted. However, with careful preliminary work it can prove useful for specific applications.

All techniques for classification improvement require that the analyst have a detailed understanding of the objects of interest and their relationships with ancillary data before attempting to improve the classification. Further, the more sophisticated the technique, the better the analyst must understand these relationships. However, in large-area, natural resource applications, these relationships are not likely to be well-known until after the classification or inventory is completed. Thus, significant improvement in large-area classification may not always be possible by these two techniques, and errors must either be tolerated, explained in the legend, or corrected through more conventional means as the inventory proceeds.

Kenk and Sondheim(1988) found that the use of ancillary data and associated prior knowledge to improve classification results of remotely sensed data during classifier operation led to a modest increase in classification accuracy. The contribution to accuracy ranged from 0.2% to 5.0%, depending on the band combination, test area, and error measure used. And they finally concluded that for increases in accuracy, the time and effort required in converting ancillary data formats and evaluating associated attributes far outweighed the returns.

A number of possible reasons for the modest performance can be considered, including inaccurate ancillary data, poor registration of ancillary data with satellite imagery, and error in developing the matrix of relationships between object classes and the ancillary classes(Kenk and Sondheim, 1988). However, the most significant factor probably relates to the distribution of the object classes in relation to the ancillary classes. Examining the ancillary *a priori* knowledge about the ground cover classes clarified the poor performance of the ancillary data. The problem lies with the inability of the ancillary data to adequately separate or differentiate the object classes within groups of object classes, in which confusion was most common. The improvement in classification accuracy as a result of the incorporation of ancillary data generally occurred in situations where definitive natural breaks in terms of the ancillary data actually exist in the landscape between the object classes.

However, the preceding exploration is not meant to suggest that ancillary data cannot play a significant role in increasing classification accuracy. This procedure may perform better with other types of ancillary data, or with similar types of ancillary data that involve either fewer or different object classes, or with ancillary data used in different procedures, such as postclassification sorting(Hutchinson, 1982). The latter would be especially effective, where well-defined naturally occurring environmental boundaries between confused classes can be identified with the ancillary data.

References

Bryant, N. A., R. G. McLeod, A. L. Zobrist, and H. B. Johnson, 1979. California Desert Resource Inventory Using Multispectral Classification of Digitally Mosaicked Landsat Frames. *Proceedings of the Symposium on Machine Processing of Remotely Sensed Data.* Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana. pp. 69~79.

Gaydos, L. and W. L. Newland, 1978. Inventory of Land Use and Land Cover of the Puget Sound Region Using Landsat Digital Data. *Journal of Research of the U. S. Geological Survey* 6 : 807~814.

Hutchinson, C. F., 1982. Techniques for Combining Landsat and Ancillary Data for Digital Classification Improvement. *Photogrammetric Engineering & Remote Sensing* 48(1) : 123~130.

Kenk, E. and M. Sondheim, 1988. Methods for Improving Accuracy of Thematic Mapper Ground Cover Classifications. *Canadian Journal of Remote Sensing* 14(1) : 17~31.

Star J. and J. Estes, 1990. *Geographic Information Systems : An Introduction.* Prentice Hall, Englewood Cliffs, New Jersey. 303 pp.