Symposium International de Cartographie Thématique
Dérivée des Images Satellitaires, Saint-Mandé 2-4 octobre 1990

○ **Extraction et qualité de l'information thématique**
Extraction and quality of thematic information

LANDUSE CLASSES DISCRIMINATION WITH SATELLITE IMAGES BASED ON SPECTRAL KNOWLEDGE

par Vladimir CERVENKA et Karel CHARVÁT

Geodetic and Cartographic Enterprise
Earth Remote Sensing Centre, Prague, Tchécoslovaquie

◆ Introduction

At present, multispectral image data are frequently exploited in many remote sensing applications. Therefore, a great attention has also been paid to the development of their automatic interpretation (classification). There are two principal approaches to the classification: supervised and unsupervised one.

A method of supervised classification of multispectral image data, based on spectral knowledge, is described in this contricution. Training data are collected for every class when using the supervised classification. The choice of training samples has to be representative, but random. However, the creation of sufficiently representative training sets may be a serious problem. Satellite images cover some hundreds km² nevertheless it is difficult to find suitable training samples, which cover the whole feature space.

It is necessary to find such classification rules, which generalize the properties of training samples (being localized in a certain part of image only) for the whole area of interest. The approach based on the creation of data base of spectral knowledge seems to be an appropriate solution of the problem mentioned. Such classification system characterizes the target classes in terms of numerical rules, which reflect characteristic relations between spectral bands. The band-to-band relations describe the shape of the spectral reflectance curves (Wharton, 1987).

◆ The spectral knowledge based classification - overview

The spectral knowledge based approach prefers the description of target classes on the basis of certain relations between individual spectral features (Wharton, 1987).This approach can be used to avoid the scene-specific limitations – the data base of spectral knowledge can be used (to a certain extent only) for classification of further scenes. The target classes are descrited using the inequalities (so called spectral relations), which can be written in the general form

$$SI > T,$$

where SI is a spectral index (a numerical expression in terms of spectral features) and T is a threshold. Every spectral relation is evaluated in regard to the ability to separate various subsets of target classes. It is necessary to find out, which target classes fulfil a certain spectral relation. This knowledge can be used for target classes discrimination using a binary tree classifier. The whole classification procedure could be divided into following steps:

- collection of suitable training samples for every class
- verification of the training sets – it is necessary to verify the labeling of training polygons and to test the separability of target classes
- editing of the training samples using k – k' nearest neighbour rule
- suggestion of appropriate spectral relations to be used for discrimination of target classes
- knowledge base development (or updating) – the results of various tests are saved in the data base
- generation of the classification algorithm
- classification of the image data.

Of course, classification method described has also some limitations:

- anomalous illumination (dense shadows) cannot be recognized correctly
- the pixels have to be relatively "pure", the spectral properties are influenced by a single target class
- the target classes can be recognized using the spectral properties only – the spatial and textural properties are not used.

◆ Verification of the training samples

The collection of a suitable training samples and the decision in which classes may by classified the satellite image data create the serious problem.

During the collection of training sets some of the training polygons are assigned to a certain class. It is necessary to verify, whether these polygons really belong to the same target class. Some methods solving the problem of unperfect labeling of training polygons (for normally distributed data) have been already investigated (Charvat, 1987). The decision are made by comparison of mean values and covariance matrices. If we do not dispose with normal data distribution, then it is possible to use a method applying mutual information (Cervenka, Charvat, 1990).

◆ Editing of training samples

It has been shown, that the editing of training set improves the performance of the classifier. The k - k' nearest neighbour method is relatively simple. The k nearest neighbours from the whole training set are found for every sample. The tested sample is edited from the training set, when not being classified in accordance with its true class membership (when at least k' of its nearest neighbours do not belong to the same class as the tested sample).

◆ The suggestion of spectral relations

The spectral relations characterize the shape of spectral reflectance curves in terms of certain inequalities to avoid the use of absolute values of individual features. The analyst can suggest an arbitrary spectral index using his empirical experience, studies of the literature or studies of spectral reflectance curves of target classes. The analyst suggests spectral indices, which seems to be typical for individual classes. Of course, the set of spectral indices from previous classification can be used. This set has been saved in the spectral knowledge data base. In this case, the values of suitable thresholds T are updated.

It is possible to use so called spectral transformations. They include linear combinations of spectral bands or various ratios of bands, e.g. brightness, greeness, yelowness (Crist, 1984) or PVI (perpendicular vegetation index (Richardson, 1977). Many authors have found out that spectral transformations are correlated with vegetation growth parameters (which can correspond to various target classes). Dusek (1985) has shown that winter wheat vegetation indices (leaf area index, percent green ground cover) are highly correlated with Thematic Mapper band ratios. For example, the three band ratios, that produce the highest correlation with the leaf area index are:

$$TM4 \,.\, TM5 \,/\, TM3$$
$$TM4 \,.\, TM5 \,/\, TM1$$
$$TM1 \,.\, TM4 \,.\, TM5 \,/\, TM2 \,.\, TM3 \,.$$

◆ The spectral knowledge base development

The aim of this operation is to find the appropriate thresholds for spectral indices so that the resulting spectral relations will have a maximal power of discrimination. The spectral knowledge based system proposes a set of possible thresholds for every spectral index in an automatic way. Then, every target class (its training samples respectively) is examined, whether (and up to which degree) fulfil the spectral relation. The results are saved into the data base. This process may be divided into five steps:

- calculation of the spectral index for all training samples
- the histogram of computed values (for every class) is created
- the threshold selection – all thresholds, which separate at least, two target classes are selected
- the test, which classes fulfil the spectral relation
- the result is saved for later evaluation.

◆ Generation of the classification algorithm

The classifier suggested is of binary tree character. A binary tree classifier assigns a class label to a sample by passing it through the tree from root to leaf. Leaves are labeled by class labels. The test of a spectral relation is performed in every node (except the leaves). Here, the features of pixel being classified are considered. If the test is fulfilled, then the test, corresponding to the right son, is performed. Otherwise, the testing continues in the left son. The label of target classes are assigned to a classified pixel at final level of the tree.

The binary tree is construted in a interactive way. The set of target classes is determined for every node (the root corresponds to the set of all classes). Then, some spectral relations are chosen from the data base which are suitable for the separation of this set of classes. Two subsets S1, S2 of target classes are distinguishable on the basis of a spectral relation, if this relation holds for subset S1 and does not hold for the subset S2. The final decision (selection of the best spectral relation for the node) is performed by the analyst. The tree classifier has several advantages, especially with multimodal data, because all features are not equally effective for the description of all target classes.

Finaly, the computer program, which realizes the classification algorithm, is generated in an automatic way. It is relatively easy, because the classifier is always of any binary tree character. All data necessary for the tree generation are available in the spectral knowledge data base.

◆ Conclusion

The method described has been successfully used at the Earth Remote Sensing Centre of the Geodetic and Cartographic Enterprise in Prague, especially for the classification of Thematic Mapper data. The main goal was to defin the land use. Several thematic map from the Northeast Bohemia region have been produced. The best classification result was approximately 92 percent accuracy over 14 target classes (using the resubstitution estimate of error). The design of binary tree classifier based on spectral knowledge is rather a time consuming process. On the other hand, the own classification of image data does not require a large amount of computing time and storage.

References

WHARTON S.W. - A Spectral-Knowledge-Based Approach for Urban Land-Cover Discrimination. IEEE Transactions on Geoscience and Remote Sensing, 1987, N° 3.

CHARVÁT K., CERVENKA V. and SOUKUP P. - Using Statistical Tests for Computation of the Classification Parameters in Remote Sensing of Earth (in Czech). In: Application of Artificial Intelligence AI: 87, ÚISK, Prague, 1987.

CERVENKA V., CHARVÁT K. and SOUKUP P. - Automatic Interpretation of TM Image Data for Defining the Land Use (in Czech). In: Application of Earth Remote Sensing Data in National Economy, CSVTS, Bratislava, 1990.

CRIST F. P. and CICONE R. C. - A Physically-Based Transformation of TM Data – the Tasseled Cap. IEEE Transaction on Geoscience and Remote Sensing, 1984, N° 3.

RICHARDSON A. J .and WIEGAND C. L. - Distinguishing Vegetation from Soil Background Information. Photogram. Eng. and Remote Sensing, 1977, pp. 1541-1552.

DUSEK D. A., JACKSON R. D. and MUSICK J. T. - Winter Wheat Vegetation Indices Calculated from Combinations of Seven Spectral Bands. Remote Sensing of Environment, 1985, pp. 255-257.

CERVENKA V. and CHARVÁT K. - Nonparametric Classification Methods in Remote Sensing. In: Application of Artificial Intelligence AI 90, ÚISK, Prague, 1990.

CERVENKA V. and CHARVÁT K. - Classification of Multispectral Imagery on Spectral Based Knowledge (in Czech). In : Digital Image Processing 89, CSVTS TESLA VUST A.S. Popova, Prague, 1989.