ANALYSIS OF THEMATIC MAPPER SATELLITE DATA FOR INVENTORYING FIELD CROPS IN NEW YORK

Xie Wei
Southwest Jiaotong University
Emei, Sichuan
The People's Republic of China
WG VII/4

ABSTRACT

A July and an August Landsat Thematic Mapper (TM) scene in Seneca County, New York state were analyzed for inventorying field crops --field corn, wheat, oats. In the analysis, enphasis was placed on digital image classification using a maximum likelihood classifier.

The results of 5 August, 4 July, 1985 TM data classification showed that accuracies of digital image classification with a maximum likelihood clssifier ranged from 91% to72% for corn, from 88% to 82% for wheat. Classification of field crops had low errors of commission. Emphasis was placed on single-date classification, however multi-date classification was also tried.

This paper also reports how to choose the best combination of fields for training data to obtain more accurate, reliable classification. Results show that choosing a different combination of training fields directly effects the classification accuracies. Some relationship between the histogram of each training field and the resulting classification accuracies are found. Analyzing the histogram of each training field before classification were very economical and eficient method to obtain accurate classification.

INTRODUCTION

Accurate and timely crop inventories provide valuable information enabling optimum use of agricultural facilities and resources. On a regional scale inventories provide information for establishing policies and for monitoring agricultural change which ultimately lead to stability in the agricultural environment. The concept of agricultural monitoring is not new, in various part of the world, including New York, some form of monitoring procedure has been established for some time. With the advent of more sophisticated remote sensing technology and data processing techniques. It became necessary to re-evaluate present established procedures in the light of current knowledge and social environment.

Many studies have demostrated the value of aircraft and satellite remote sensing for crop identification and inventory. Especially TM data makes more attractive for field crop inventory and monitoring.

The purpose of research reported here was field crop identification in the heterogenous region of New York state. Present census techniques are based on extensive field work and questionnaires. The procedure is large and cumbersome requiring extensive man-hours and good co-ordination. It is difficult to provide accurate and timely information, because of the dispersed nature of the field crop. This studying was undertaken to determine the extent to which Landsat TM data, might provide useful information for the census of field crops in New York state.

STUDY AREA AND MATERIALS

The study site used in this analysis is in Seneca County, New York state, located northeastern of Seneca Lake. The major land cover classes in this area are water, corn, wheat, clover, oats. TM images of 4 July, 1985 and 5 August, 1985 were chosen. The TM scene is identified by path 16/row 30 in the Worldwide Reference System.These images represented different stage of the same growing season. They were little cloud and of good quality.Grond surveys of the study site were conducted by CLEARS that summer to identify the field crops in that area. There were 10 fields of field corn (each field was represented by a certain number, No1, No.6, No.7, No.8, No.9, No.10, No.11, No.13, No.15, No.21), 5 fields of wheat (No.3, No.4, No.12, No.14, No.14-1), 1 field of clover (No., 5) and 1 field of oats (No.20). Corn was planted during the period from mid-May to mid-June in 1985. It grew to knee height by the beginning of July. However, it did not have significant ground cover until mid-July. The real growth of biomass was from mid-July, to mid-August, which was reflected by the increase in intensity of reflectance of the infrared band. Oats were planted as early as mid-April and harvested in August. Winter wheat was planted from September to October and harvested is August.

The hardware and software used for the image analysis reside at the Cornell's School of Civil and Environmental Engineering. The heart of this facility is International Digital Image Systems mode 70, linked to VAX 11/750 minicomputer.

VISUAL ANALYSIS

The TM date for July and August, 1985 were first analysed visually on IIS display. The various field crops and different field of the same crop were studied to determine how well field crops could be identified.
Different band combination were examined. The best results were found displaying bands 3, 4,and 5, in blue, red, and green respectively. Field crops could often be identified on the IIS display by their color and shape. At different growning season, the field crops have different color in the IIS display. For example, the corn have green image in the July image, but red image in August image in the same bands conbination. The field corn were generally identified, but confused with oats. The one field of oats had the same reflectance characteristics as corn in all seven bands TM data, but oats were planted as early as mid-April. So it would be possible to detect the greenness of oats fields in June. The wheat was easily identified.

DIGITAL ANALYSIS

In order to assess the spectral properties of field crops in the seven TM bands, the digital counts of each crop field were examined. The means and standard deviation of the TM digital count of each field studied are shown in Tables 1 and 2. In July digital counts of most fields of the same crop vary in the same bands. For example, in band 1 the digital counts of field corn vary from 105 to 136, but in August, the digital counts of the same fields vary from 115 to 127. The standard deviation of TM digital counts of each corn field in July is more than the standard diviation of TM digital counts of each corn field in August. This means the August TM image digital counts are more uniform than the July image digital counts. Because the field corn was young in July with low percent covers. The reflectance of the field corn included reflectance of different soil backgrounds. In August, however, most field corn in the region were mature with relatively complete crop cover regardless of planting date. Comparing the digital counts of different field crops. The field crops were separable.

CLASSIFICATION

Classification was done using a supervised maximum likelihood classifier, with a decision threshold of two standard deviations. July 4 and August 5, 1985 subscenes were each classified with different combinations of the 7 bands to determine the best bands for classification. Training was done using different combinations of the known fields to determine the best combination of training areas for high classification accuracies and low errors of commission. The results are shown in the confusion matrices(Table 3 and 4). In August, corn and wheat were classified with 91% and 86% correct classifications, respectively. In July, corn and wheat were classified with 80% and 86% correct classifications, respectivly.

More accurate classifications are obtained when the training data meet the assumptions of the classifier (Story et al, 1986).Hixson et al, (1980) concluded that the major variable affecting classification accuracy was the training method used to calculate the class statistics. They went on to say that all cover types should be adequately represented by a sufficient member of samples in each spectral class, thus guaranteeing a true representation of the spectral charactristics of the class. Key to the success of digital classification are representive fields for training and testing (Williams, 1986). So how to choose the training area is very important. In this study we found some relationship between the histogram of each training field and the classification accuracies.

Examination of histogram is useful and often necessary, preliminary step for successful classification of field crops. Ideally, class histograms tend to be Gaussian distributions. After measuring the histogram of every crop fields known. We found that histogram shape varied not only different crops but also within the same crop. Some histograms had multiple peack values and some had an extended tail toward higher or lower grey levels. These different shapes are due to the different planting date, management practices, soil condition and other environmental factors. After analyzing the histograms of the field crops, we found that when the histogram more closely approach a normalized Gaussian distribution, the classification becomes more accurate. Coversely, when the histogram of crop fields have several peack values and extended tail, the classification accuracy was lower with higher errors of commission. When the pixels are less certain limitation values (i. e. the field too small) the histogram of that training site is a separate distribution. It is impossible to get accuracy classification using these fields for training data. Becouse the separate histogram of crop field did not satisfied the assuption of the maximum likelihood classifier.

Given a certain time and certain environment, the histogram of every kind of crop has certain shape. It is difficult to know the actural shapes. However, we can use statistical method to obtain its approximate shape. For example, the histogram of No.1 and No.8 corn fields are the same as the histograms of No.1, No.6, No.8, No.11, No.13 and all ten corn fields. increasing the fields combination, their histogram does not change. We call this histogram as total histogram of corn in this time and this evironment. It contains all histograms of every corn field, i.e. every histogram of corn field is overlaped in the total histogram.

In general, the more pixels and fields of crop are used as a training data, the higher classification accuracies are. But up to certain degree, the classification accuracies do not increased with the increasing of pixels and fields. For example, using nine corn fields and using five corn fields as a training data, the classification accuracies are almost the same.

How can I evaluate the reliability of classification? Generally, if the fields chosen for training data are correct, in spite of using any fields as a testing data, the classification accuracies do not change. It means that these fields are well representive of the crop. When using some fields for testing data, the classification accuracies are high, but when using other fields for testing data, the classification accuracies are low These classification accuracies are not reliable.

How many fields chosen as training data are just adequate? There are many choices you can make. Different fields combination give different results. The best way to choose the training histograms of crop fields is to choose the fields with histogram which fill the total histogram of this crop. Such as we randomly choose band 5 in August TM data, the histogram of the No.1 and No.8 corn fields just fill the total histogram of corn. The histogram of No.1 corn field lie in the right of total histogram, the histogram of No.8 corn field lie in the left of the total histogram. otherwise, any other field combination which histograms filling the total histogram of corn, must be more than the two fields. Such as we choose No.1 No.6 No.8 No.9 No.13 corn fields for training data. The histogram of No.6 No.9 No.13 corn fields just overlap with the histograms of No.1 and No.8 corn fields. The No.6 No.9 No.13 corn fields can be omitted from the training data, just using No.1 and No.8 corn fields for training data is enough. The classification accuracies are the same. So choosing the No.1 and No.8 corn fields for training data are the economical representative of all corn field for classification. If you did not analysis the histograms of every field before classification, you have to try many times classification to find out the best field combination. It waste the CPU time. Here just analyzing the histogram of band 5, the more bands are analysis, the better.

CONCLUSION

Using visual and digital image analysis of a July and an August Landsat Thematic Mapper scene, wheat and corn in New York can be identified automatically.
The best accuracies of digital image classification with a maximum likelihood classifier are 91% for corn 86% for wheat.

This study has shown that the classification accuracies are dependent upon the inherent characteristics of the class and the choice of fields were as training data. Analysis of the histogram of each field of the crop before classification results in broader signature. Chossing the histogram of the combined fields as training data fit the total histogram of this class. The combination fields will adequately represented this class. The classification accuracies will be higher and more reliable. With any inadequate combination of the fields as training data, the classification accuracies are not reliable.

ACKNOWLEDGMENT

The author would like to thank CLEARS of the Cornell's School of Civil and Environmental Engineering for Providing equipments and meterials for this project. The author thanks Professor Ta Liang, Warren R. Philipson, William D. Philpot for assistance and helpful suggestions in the preparation of this manuscript and Laurie Trolier, Sally Buechel for the drafting and valuable disscussing.

REFRENCES

Hixson, M.M., D. Scholz, N.Fuhs, and T. Akiyamal,1980, Evaluation of Several Schemes for Classification of Remotely Sensed Data photogrammetric Engineering and Remote Sensing. 46;12;1547-1553.

Story, M.H., J.B. Compbell,1986, The Effect of Training Data Variability on Classification Accuracy Photogrammetry 1986, ACSM -ASPRS Annual Convention Vol4, p370-378.

Philipson W.R., T.Liang, 1975. Airphoto Analysis in the Topics: Crop Identification. Proc. of the Tenth Int. Symp. on remote sensing of the Environment. October 6-10. ERIM Ann Arbor, Michiban.

Philipson W.R., V.L. Williams, D.K. Gordon, and W.D. Philpot, Vegetable and Fruit Inventory with Landsat TM Data: Preliminary Results 1985 ASP-ACSM Convention vol.1 p39-46

Thomas H.C. Use of Multitemporal Spectral Profiles in Agricultural Land-Cover classification. Photigrammetric Engineering and Remote Sensing April, 1986 p535-544

Williams, V.L., 1986 Analysis of Thematic Mapper Satellite Data for Inventorying Commercial Vegetable Crops in New Toyk. Unpublished masters thesis presented to Cornell University.

TABLE 1 TM DIGITAL COUNTS OF FIELDS 5 AUGUST 1985 (MEAN VALUE AND STANDARD DEVIATION)

Field	crop	band 1	band 2	band 3	band 4	band 5	band 6	band 7	pixels
No. 7	corn	115(2)	43(1)	42(1)	131(5)	78(3)	135(1)	22(2)	176
No. 13	corn	127(8)	52(6)	60(12)	97(5)	105(18)	141(4)	50(17)	321
No. 3	wheat	123(3)	49(2)	57(4)	88(4)	127(6)	146(2)	127(6)	193
No. 4	wheat	132(6)	56(4)	71(10)	94(5)	154(16)	150(7)	74(12)	41
No. 20	oats	116(3)	44(2)	43(4)	137(8)	84(9)	136(1)	25(8)	201
No. 5	clover	119(4)	46(2)	48(4)	101(8)	117(10)	147(2)	43(6)	221

TABLE 2 TM DIGITAL COUNTS OF FIELDS 4 JULY 1985 (MEAN VALUE AND STANDARD DEVIATION)

Field	crop	band 1	band 2	band 3	band 4	band 5	band 6	band 7	pixels
No. 7	corn	105(4)	45(3)	51(6)	94(7)	107(8)	147(1)	50(7)	185
No. 13	corn	136(13)	64(8)	86(16)	90(10)	143(17)	153(4)	86(12)	324
No. 3	wheat	105(6)	47(5)	58(8)	99(8)	96(9)	142(2)	35(5)	172
No. 4	wheat	134(6)	63(5)	89(9)	86(6)	141(9)	146(2)	87(7)	96
No. 20	oats	113(6)	50(4)	62(9)	100(6)	131(12)	146(2)	67(9)	194
No. 5	clover	98(13)	39(9)	39(13)	136(11)	97(12)	142(3)	32(9)	299

TABLE 3 Confusion Matrix for August 5 scene, band 1-7(threshold=2)

	Percent of classified pixels				
	corn	wheat	clover	uncl	pixels
corn	91.3	0.4	0.1	8.5	951
wheat	00	86.1	1.5	12.4	461
colver	00	6.9	83.7	9.4	277

TABLE 4 Confusion Matrix for July 4 scene, band 1-7(threshold=2)

	Percent of classified pixels				
	corn	wheat	clover	uncl	pixels
corn	80.7	4.6	7.1	7.6	1001
wheat	6.9	86.2	4.5	2.5	406
colver	8.0	4.8	83.9	3.2	311