N78-14539

# CLASSIFICATION OF LANDSAT AGRICULTURAL DATA BASED UPON COLOR TRENDS

J. D. Tubbs

Department of Mathematics
University of Arkansas
Fayetteville, Arkansas

## SUMMARY

Historically, supervised classification procedures are usually "trained" by photo-interpreters who preprocess the data. This preprocessing consists of identifying and labeling training areas by analysing photo products produced from the digital Lansat data. An oversimplification of this procedure is that the photo-interpreters label a field according to whether or not its observed color (determined by the false color IR film product produced from a composition of landsat bands 4, 5, and 7) coincides with the expected color of known crops at acquisition time t. Since the success of most, if not all, supervised classification procedures is highly dependent upon the success of the photo-interpreters, multiple acquisitions obtained throughout the growing season are desired in order to insure sufficient confidence in the field labeling. In fact, very few errors are encountered if the photo-interpreter has imagery available at the various growth stages of the crops he is interested in identifying. Based upon this assumption, it seemed feasible to automate a part of the photo-interpreters logic process, that of labeling or classifying fields (hence pixels) according to their color trend. The purpose of this paper is to present such a classification procedure. The decision rules have been developed for classifying an unknown observation by matching its color trend with that of expected trends for known crops. The results of this procedure have been found to be encouraging when compared with the usual supervised classification procedures.

## 1. INTRODUCTION

Historically, the majority of crop inventoring of agricultural regions using Landsat multispectral scanner data have been performed using supervised classification procedures [MacDonald 1975]. This type of procedure usually consists of having trained photo-interpreters (PI) define and label homogenous areas found on a film product produced from the digital data. A principle product used in this identification is a false color IR produced from a composite of Landsat bands 4, 5, and 7. In defining and labeling these homogenous areas, the PI selects areas which are representative of the major crops to be classified in the test segment. Since the performance of the classifier is highly dependent upon its training inputs (hence the PI labeling), it is often desirable and necessary for the PI to analyze multiple acquisitions of the same test area throughout the growing season. An oversimplification of this procedure is that the PI labels a field, crop x, according to whether or not its color at acquisition time t coincides with the expected color for crop x at time t. By repeating this process for each acquisition, the PI obtains a degree of confidence in the labeling of the training inputs. In this paper, I have developed a simple procedure which attempts to automate this portion of the PI logic process, that of labeling fields or pixels according to their observed color trend. The results of this procedure are compared with a supervised classification procedure, both in terms of spacial properties (preservation of field structure) and as crop proportion estimates. Due to the nature of the available data, I have restricted my attention to the problem of separating wheat from all non-wheat by using Landsat imagery obtained from at least three distinct growth stages for wheat.

## 2. PROCEDURE

Engvall, Tubbs, and Holmes (1977) have shown that certain agricultural crops could be classified according to their temporal trend. In order to detect the temporal trends, it was necessary to transform the original four dimensional Landsat space $S_L$ onto a two dimensional space $S_D$. Then an unknown observation is classified by matching its temporal trend with the temporal trend for known crops. In their paper, the transformation was based upon band differencing and ratioing, in order to take advantage of the spectral variability over varying ground targets. In this paper, I have developed a procedure which is also based upon classifying temporal trends, however I have projected the original Landsat space $S_L$ onto a hopefully more familiary space, one that is equivalent to plotting points in the standard CIE Chromaticity diagram [ 3 ].

As mentioned in the introduction, the analysis of false color IR film product obtained periodically throughout the growing season is usually sufficient to insure that the PI correctly label the training inputs. This is particularly true if certain auxillary information (cropping practices, updated weather reports, etc.) is also available. The PI then labels a field A, crop x, if the observed color (red on film implies a green-like target, non-red implies a non-green like target) agees with the expected color for crop x at acquisition time t. For example, if the PI suspects field A is winter wheat, then he expects to see a non-red (preemergence), red (emergence to heading), then non-red (harvest). There will be some variability in this general trend, however variance could be anticipated according to the auxillary information. Since the film product is produced directly from the digital data, it seems feasible that one could determine directly (without analyzing the film product) whether or not an observation is red or non-red. Then by repeating this procedure for each acquisition, one would be able to classify an unknown observation according to its agreement or disagreement with the expected trends for known crops.

Mathematically this is equivalent to defining a transformation from the spectral space $S_L$ onto 0 to 1 (0 = non-red, 1 = red). Let $X_t=(x_{1_t},x_{2_t},x_{3_t},x_{4_t})$ represent the four dimensional row vector corresponding to the multispectral scanner data ($X_t \varepsilon S_L$) acquired at time t. Due to the lack of variability in the scanner data and to the inconsistency in the dynamic range of the spectral bands, it is often necessary to enhance the data before processing the color product. One such approach is to define a linear map on $X_t$ by

$$x^*_{j_t} = s_j x_{j_t} + b_j \quad \text{for } j = 1,2,3,4. \tag{1}$$

where $s_j$, $b_j$ are such that $\bar{x}_j-3sd_j=0$ and $\bar{x}_j+3sd_j=255$, when $\bar{x}_j$, $sd_j$ are the sample mean and standard deviation for band j and (0,255) represents the dynamic range of the photo processor. Now define the following transformation

$$Y_t + A\,X^*_t$$

$$\text{where } X^*_t = (x^*_{1_t},x^*_{2_t},x^*_{3_t},x^*_{4_t}),\ br = x^*_{1_t} + x^*_{2_t} + x^*_{4_t} \text{ and} \tag{2}$$

$$A = \begin{matrix} 0 & 0 & 0 & 1/br \\ 0 & 1/br & 0 & 0 \\ 1/br & 0 & 0 & 0 \end{matrix}$$

Thus $X^*_t$ is mapped onto $Y_t= (y_{1_t},y_{2_t},y_{3_t})$ which is equivalent to $Y_t=(y_{1_t},y_{2_t},br)$ since $y_{1_t}+y_{2_t}+y_{3_t}=1$. Note that $Y_t$ is the Trichromatic coefficient [ 3 ], and $y_{1_t} y_{2_t}$ are the chromaticity coordinates which can be plotted on the CIE diagram. These coordinates can be used to determine whether or not $X_t$ is red or non-red when projected onto the color plane $S_C$. Let R denote the red region of the color plane and NR the non-red region. If $(y_{1_t}, y_{2_t}) \varepsilon$ R, then let $d_t = 1$, otherwise $d_t = 0$. By repeating this process for each acquisition t, t=1,2,..., n, we have reduced a 4 x n dimensional row vector $X = (X_1, X_2, \ldots, X_n)$ onto a n dimensional row vector $D = (d_1, d_2, \ldots, d_n)$, where $d_j = 1$ if the $j^{th}$ acquisition of the unknown observation X is red like on the film product.

## 3. CLASSIFICATION PROCEDURE.

Suppose that there are m different crops to be classified, denoted by $C_1, C_2, \ldots, C_m$. Let $G_j$= the set of $D_k$'s that are to be associated with crop j, j=1,2,...,m. Let $X = (X_1, X_2, \ldots, X_n)$ be an arbitrary observation to be classified. Since there are $2^n$ possible outcomes and n is usually small ($n \leq 6$) the matching problem is fairly trivial. For example, assume that n=4, where the acquisitions are spaced according to the distinct growth stages for wheat. Let $C_1$ denote the class of pure wheat, $C_2$ denote the class of possible wheat, and $C_3$ denote the class of non-wheat, where $G_1 = \{D_1\}$, $G_2 = \{D_2, D_3, D_4\}$ and $G_3$= the remaining 12 outcomes and

| | | |
|---|---|---|
| $D_1 = (0,1,1,0)$ | ideal wheat (z=6) |
| $D_2 = (0,1,0,0)$ | possibly early maturing wheat (z=2) |
| $D_3 = (0,0,1,0)$ | possibly late developing wheat (z=4) |
| $D_4 = (1,1,1,0)$ | possibly very early developing wheat (z=7). |

Class $C_2$ is defined to account for the variability of the wheat fields within a sample segment and the proportion of observations falling into this class could be weighted according to the nature of the auxillary information. Since the D vector is made up of zeroes or ones, there is an easy procedure for matching unknown observations with the $2^n$ possible outcomes. Let $z = d_1 + 2d_2 + 4d_3 + 2^{n-1}d_n$. Now the comparison of the color response for an unknown observation with that of known crop consists of comparing integers. That is, in our example if the z value for the unknown observation was a 6, then it was classified as pure wheat, if z = 2,4,7 then it was classified as possible wheat and would be weighted according to the available information, and if the z value was not 2,4,6,7 it was classified as non-wheat.

## 4. RESULTS

The proposed classifier (color) was applied to 19 different Landsat data sets collected during the 1974-1975 growing season throughout the major wheat producing regions of the United States. To aide in the evaluation of this procedure, the 19 data sets were classified using three different procedures. The first (MLE) is maximum likelihood classification using any or all of the available Landsat acquisitions. The results were deemed satisfactory by the PI using ad-hoc investigation. The second procedure (B104) is also based upon maximum likelihood classification where the data sets were classified using all available acquisitions (multitemporal classification). The third procedure is the non-supervised classification procedure (DELTA) as proposed by Engvall, et al. (1977). The results (wheat proportion estimates) have been summarized in Table 1. Table 2 lists correlations and linear regression coefficients.

ORIGINAL PAGE IS
OF POOR QUALITY

Table 1: Wheat Proportion Estimates

| Data Set | Procedures | | | |
|---|---|---|---|---|
| | COLOR | MLE | B104 | DELTA |
| 1 | 11.9 | 18.0 | 12.6 | 6.0 |
| 2 | 15.9 | 26.0 | 35.2 | ** |
| 3 | 20.0 | 7.0 | 10.9 | 27.5 |
| 4 | 7.7 | 2.0 | 7.4 | 15.9 |
| 5 | 14.2 | 14.6 | 18.7 | 11.3 |
| 6 | 22.6 | 36.9 | 33.1 | 24.2 |
| 7 | 20.7 | 8.0 | 13.8 | 19.1 |
| 8 | 9.1 | 15.3 | 15.3 | 34.5 |
| 9 | 1.6 | 0 | .2 | 7.0 |
| 10 | 32.1 | 21.3 | 23.2 | 29.7 |
| 11 | 15.1 | 12.7 | 2.4 | 8.7 |
| 12 | 18.9 | .3 | 5.3 | 3.4 |
| 13 | 34.9 | 40.0 | 39.5 | 45.8 |
| 14 | 9.6 | 8.5 | 8.5 | 5.1 |
| 15 | 16.0 | 8.8 | 7.7 | .5 |
| 16 | 9.4 | 13.7 | 13.7 | 1.5 |
| 17 | 23.1 | 46.7 | 12.3 | 54.6 |
| 18 | 3.4 | 20.0 | 19.3 | 7.4 |
| 19 | 26.6 | 18.0 | 15.2 | 7.0 |
| $\overline{X}$ | 16.1 | 16.7 | 15.4 | 17.1 |
| 5D | 9.0 | 13.0 | 10.7 | 15.7 |

** Dates were unsatisfactory

Table 2: Correlation and Regression Coefficients

| Source | $B_o$ | $B_1$ | $R^2$ | r |
|---|---|---|---|---|
| Color x MLE | 8.14 | .47 | .46 | .68 |
| Color x B104 | 10.00 | .35 | .35 | .64 |
| Color x DELTA | 7.69 | .54 | .41 | .59 |

## 5. CONCLUSION

As mentioned in Engvall, et al (1977), the difficulty in evaluating or selecting "best" classification results in absence of ground truth information has been a real problem to remote sensing investigators. This is particularly true if one wishes to make decisions based solely upon fixed decision rules. Recently, there has been a trend toward developing simple (minimum setup and computer time) classification procedures to aid in this decision process. In this paper, I have introduced such a procedure. One of the main advantages of this procedure is that the transformation is similar to the one used in producing the color film product. This is particularly important when one considers the nature of this transformation. The transformation is many to one, which means that different points in the spectral space are mapped to the same color. Mathematically this is equivalent to saying that the inverse transformation is not unique. This is evident when the PI defines two training areas as identical crops, only to find that the classifier finds that they are entirely different. This problem is particularly apparent in defining representative multitemporal training areas. For this reason I wanted to classify a sample segment using the same space in which the training areas were defined and labeled. It was my hope that the color process (hence the proposed classifier) could be used to separate major crop types (wheat from non-wheat), whereas it may not be able to distinguish different varieties of wheat.

The preliminary results of this investigation have been encouraging. There seems to be a very good agreement between the classification results and the PI labeled training areas. I stated in the introduction that I wanted to compare this procedure with some of the usual procedures using both spacial properties and proportion estimates. Due to lack of resources, I was not able to display any of the class maps generated using this procedure, however they were very satisfactory (comparison of computer printer plot with 5 x 6 color film product). The proportion estimates also seemed satisfactory, although one must always guess at what is a satisfactory answer.

## REFERENCES

(1) Engvall J.L., Tubbs J.D., and Holmes Q.A. (1977), "Pattern Recognition of Landsat Data Based Upon Temporal Trend Analysis" to appear in Journal of Remote Sensing of the Environment.

(2) MacDonald R.B. et al (1975), "The Use of Landsat Data in Large Area Crop Inventory Experiment (LACIE)," Proceedings of the Symposium on Machine Processing of Remotely Sensed Data, West Lafayette, Ind. pp. IB-(1-23).

(3) Wenderoth, et al (1972) "Multispectral Photography for Earth Resources" Science Engineering Research Group Long Island University. Greenvale, New York.