1-1-1975

Signature Extension Using the MASC Algorithm

R. G. Henderson

Henderson, R. G., "Signature Extension Using the MASC Algorithm" (1975). *LARS Symposia.* Paper 44.
http://docs.lib.purdue.edu/lars_symp/44

Reprinted from

Symposium on

Machine Processing of

Remotely Sensed Data

June 3 - 5, 1975

The Laboratory for Applications of
Remote Sensing

Purdue University
West Lafayette
Indiana

IEEE Catalog No.
75CH1009-0 -C

Copyright © 1975 IEEE
The Institute of Electrical and Electronics Engineers, Inc.

SIGNATURE EXTENSION
USING THE MASC ALGORITHM

R. G. Henderson

Environmental Research Institute of Michigan
Ann Arbor, Michigan

I. ABSTRACT

A new signature extension method for use with LANDSAT data has been developed. The MASC (Multiplicative and Additive Signature Correction) algorithm uses an unsupervised clustering routine to gain relative information from two data sets. This information is then used to map the signatures derived from one data set onto the other data set. The MASC algorithm can be totally automated, thus making it suitable for use in large area crop inventories.

This signature extension method has been tested on agricultural LANDSAT data. The results of field center pixel classification using MASC-extended signatures have been compared with classification results using untransformed signatures. In all three data set pairs the MASC algorithm yielded very good results.

II. INTRODUCTION

With the development of satellite multispectral scanners (MSS) it has become easier to gather data from large areas. This data collection effort has the potential of providing timely information concerning the state of the world-wide crop production. In order that this potential is realized it is necessary to find methods of processing the data in a timely and cost effective manner.

A major stumbling block in the way of achieving cost effective processing is the requirement of large amounts of ground information. This ground information is required to train the computer to recognize different crop types. Because of variations in measurement conditions when the data is collected, the crop signatures are not constant in either time or space and the computer must be re-trained on a regular basis in order to prevent the degradation of recognition accuracy. The need to retrain the computer requires the acquisition of new ground information which is both costly and time consuming.

One objective of signature extension is the mapping of target spectral information (signatures) from a training data set (TDS) to a spatially-temporally removed recognition data set (RDS). If this can be successfully accomplished then one set of ground information, from the TDS, can be used in the processing of several data sets.

III. SOURCES OF DATA VARIABILITY

There are a number of factors which can be sources of variation in scanner signals. Some of these sources are listed below, where we have divided them into three categories: instrumental sources, environmental sources, and scene related sources of variation.

SOURCES OF VARIATION IN MULTISPECTRAL SCANNER SIGNALS

A. Instrument
  Scanner Electronics and recorder instabilities
  Gain changes
  Nonuniform angular responsivity

B. Environment
  Changes in irradiance
  Changes in atmospheric transmittance
  Changes in atmospheric path radiance

C. Scene
  Geometric effects
  Reflectance effects

Instrumental sources are associated with the mechanics, optics, and electronics of the multispectral scanner. Included in this category are gain changes, non-uniform angular responsivity, and other recorder and electronic instabilities. Since many of these effects are deterministic, they can be eliminated from the data during an initial data preparation stage.

Environmental sources of variation include changes in the magnitude and spectral make-up of the irradiance at ground level, changes in atmospheric transmittance, and changes in path radiance. Changes in irradiance result from changes in the atmospheric state as well as from solar positional changes that occur between the times the data sets are collected.

Atmospheric transmittance and path radiance will also change as the atmospheric state changes. In Fig. 1 we see the variation of path radiance, as calculated using the ERIM Radiative Transfer Model [Turner, 1973], for different atmospheres (as represented by visibility). It is clear that path radiance can vary considerably with changing atmospheric state, up to 37% for the visibilities shown in Fig. 1. These quantities are also

functions of scan angle since they depend on the path length from the ground to the scanner. Shown in Table 1 are calculations of the effect of scan angle on both the path radiance and total radiance received by the LANDSAT scanner. The change in path radiance over a range of scanner view angles, from +6° to -6° relative to nadir, is greater than 18%. The change in total radiance, over the same range of view angles, can be as large as 10%.

The in-scene effects are of two types. The first effect is the geometric variations due to sun-angle and bidirectional reflectance. These will cause the amount of radiation reflected in a particular direction to depend on time of day and position of the target in the scene. The other in-scene effect is variation in target reflectance. This may be caused by differences in moisture content of the soil or soil type. Also differences in irrigation and fertilization or crop vigor will cause variation in the crop reflectances.

IV. PHYSICAL BASIS OF MASC ALGORITHMS

To see how these effects combine to affect the variability of the MSS signals we write the equation for the signal recorded by the scanner in channel i for crop α,

$$S_\alpha^{(i)} = G^{(i)}E^{(i)}T^{(i)}\rho_\alpha^{(i)} + G^{(i)}L_p^{(i)}. \quad (1)$$

The instrumental effects are contained in the gain term $G^{(i)}$, while the atmospheric effects are contained in the irradiance $E^{(i)}$, the transmittance $T^{(i)}$, and the path radiance $L_p^{(i)}$. The in-scene effects are contained in the reflectance $\rho_\alpha^{(i)}$. Thus, the effect of variations in atmosphere and instrumental response is to produce both multiplicative and additive variations in the recorded signal. The in-scene variations will produce multiplicative variations in the scanner signal.

The signals actually recorded for the same crop from two different data sets are:

$$S_{\alpha 1}^{(i)} = G_1^{(i)}L_{\alpha 1}^{(i)} = G_1^{(i)}E_1^{(i)}T_1^{(i)}\rho_{\alpha 1}^{(i)} + G_1^{(i)}L_{p1}^{(i)}, \quad (2)$$

$$S_{\alpha 2}^{(i)} = G_2^{(i)}L_{\alpha 2}^{(i)} = G_2^{(i)}E_2^{(i)}T_2^{(i)}\rho_{\alpha 2}^{(i)} + G_2^{(i)}L_{p2}^{(i)}, \quad (3)$$

If we wish to extend the signatures extracted from data set 1 (the TDS) to data set 2 (the RDS), in a way which will yield accurate recognition, then we must find a mapping such that

$$S_{\alpha 2}^{(i)} = A_\alpha^{(i)} S_{\alpha 1}^{(i)} + B_\alpha^{(i)}. \quad (4)$$

By substituting equations (2) and (3) into equation (4) it is found that

$$A_\alpha^{(i)} = \frac{G_2^{(i)}E_2^{(i)}T_2^{(i)}\rho_{\alpha 2}^{(i)}}{G_1^{(i)}E_1^{(i)}T_1^{(i)}\rho_{\alpha 1}^{(i)}}, \quad (5)$$

and

$$B_\alpha^{(i)} = G_2^{(i)}L_{p2}^{(i)} - A_\alpha^{(i)}G_1^{(i)}L_{p1}^{(i)}. \quad (6)$$

Equation (4) defines a multiplicative and additive signature correction (MASC) which maps the signature for crop α in the TDS onto the signature for crop α in the RDS. What is necessary for successful signature extension is to obtain the MASC parameters $A_\alpha^{(i)}$ and $B_\alpha^{(i)}$.

The two parameters $A_\alpha^{(i)}$ and $B_\alpha^{(i)}$ contain the effects of all measurement variables including target reflectance. If the distribution of reflectances for target α is different for the two data sets then the MASC mapping will, in general, not be unique. The two MASC parameters will have an explicit dependence on the target type α. In what follows we will assume that the distributions of reflectances for the various targets are approximately the same for the two data sets. In this way we are able to employ a unique mapping using the parameters $A^{(i)}$ and $B^{(i)}$. If the above assumption does not hold then we will define a unique mapping by the parameters $A^{(i)}$ and $B^{(i)}$ which are the averages over α of the parameters $A_\alpha^{(i)}$ and $B_\alpha^{(i)}$,

$$A^{(i)} \equiv \sum_\alpha a_\alpha^{(i)}A_\alpha^{(i)}$$

$$B^{(i)} \equiv \sum_\alpha b_\alpha^{(i)}B_\alpha^{(i)}$$

where

$$\sum_\alpha a_\alpha^{(i)} = 1$$

$$\sum_\alpha b_\alpha^{(i)} = 1$$

Thus equation (4) becomes

$$S_{\alpha 2}^{(i)} = A^{(i)}S_{\alpha 1}^{(i)} + B^{(i)}. \quad (7)$$

V. THE MASC ALGORITHM

So far everything we have done is formal and of little use unless the MASC parameters can be found for the data sets of interest. If the gain and target reflectances were the same for both data sets, $A^{(i)}$ and $B^{(i)}$ could in theory be obtained by making appropriate atmospheric measurements at the time of data collection. Even this, however, may not be practical for timely large area inventories.

What is required is equivalent "looks" at the two data sets. In this way information concerning the relative natures of the data sets can be obtained without resort to ground observations. One method of obtaining this information quantitatively is with the use of unsupervised clustering. The MASC algorithm which has been developed to obtain $A^{(i)}$ and $B^{(i)}$ uses an ERIM clustering routine. Any good clustering routine should work provided it be applied in exactly the same way to both data sets.

The clustering routine is applied separately to both data sets. (It isn't necessary to cluster over every point in the data set, a sampling, e.g., over every other scan line would be sufficient.) The output from the ERIM clustering routine is a set of clusters. The number of pixels in each cluster is given in the output. The clusters are represented by multivariate Gaussian distributions. Only those clusters are retained which contain more than 1% of all the pixels clustered.

These clusters are unidentified for both data sets; no ground truth has been used. In order to use these clusters to obtain the MASC parameters of equation (7) it is necessary to find a correspondence between the individual clusters of each data set. To form this correspondence we order the clusters of each data set on the basis of their means in one of the channels. Other, perhaps better, methods of forming this correspondence are in the process of being programmed and tested. In the implementation used on the data reported here the channel chosen for this ordering is the channel with the largest range of values. After both sets of clusters have been ordered in this way a one to one correspondence is made -- the number one cluster of data set one is matched up with the number one cluster of data set two, etc. Using the means of the Gaussian distributions representing the clusters as points defining a line

$$c_2^{(i)} = \hat{A}^{(i)} c_1^{(i)} + \hat{B}^{(i)} \tag{8}$$

where the $c_2^{(i)}$ and $c_1^{(i)}$ are the set of cluster means in channel i for data set 2 and data set 1, respectively. A regression routine is used to give a straight line fit to equation (8). Any pair of clusters whose percentage deviation from the line is more than 10%, for any channel, is eliminated from the cluster sets and regression is reentered. This editing is done to try to minimize the number of cluster pairs which have been incorrectly matched. The parameters $\hat{A}^{(i)}$ and $\hat{B}^{(i)}$ which result from this second regression are then applied to the signatures of the TDS, as in equation (7). The transformed signatures are then applied to the RDS.

VI. RESULTS

The MASC algorithm has been tested on three LANDSAT data set pairs [Henderson, 1975]. These were from the CITARS (Crop Identification Technology Assessment for Remote Sensing) study [Malila, 1975(a); Hall, 1974; Malila, 1975(b)]. The data sets used for training were from Fayette Co., Illinois, June 11, 1973 (Fay 6-11) and White Co., Indiana, August 31, 1973 (White 8-21). The signatures from Fay 6-11 were extended to the recognition data sets from Fayette Co.,Illinois, June 10, 1973 (Fay 6-10) and Shelby Co., Indiana, June 8, 1973 (She 6-8). The White 8-21 signatures were extended to the Fayette Co., Illinois, August 21, 1973 (Fay 8-21) data set. During the June period wheat was considered the major crop and five signatures were used for purposes of recognition: wheat, wheat 2, water, trees, bare soil,and weeds. For the August period the major crops were corn and soybeans. The signatures used were: corn, soybeans, pasture, quarry, and trees.

Both MASC transformed and untransformed signatures from the TDS were applied to the RDS using the ERIM Linear Classifier [Crane, 1973]. The results of field-center pixel recognition are listed in Table 1. For the June period the results are tabulated as the percentage of wheat pixels that were recognized as wheat (Correct Wheat) and the percentage of non-wheat pixels that were recognized as anything other than wheat (Correct Other). For the August data sets the results are tablulated as the percentage of corn pixels recognized as corn (Correct Corn), the percentage of soybean pixels recognized as soybeans (Correct Soy), and the percentage of non-corn, non-soybean pixels that were recognized as anything other than corn or soybean (Correct Other).

As can be seen in Table 2, the MASC transformed signatures performed significantly better than the untransformed signatures. This can be seen even more clearly in Table 3 where we list the average probabilities of misclassification (based on Table 2) for our three data set pairs. It is also of interest to note the relative stability of the probability of misclassification for the MASC signatures in comparison to the probabilities when untransformed signatures are used. Because of the essentially random nature of the variations between data sets it is to be expected that the use of untransformed signatures will yield random recognition accuracies. If, however, a signature extension algorithm is used which is able to correct for the variations between data sets then one would expect that the recognition accuracy should be relatively constant. This appears to be true for MASC as is evident in Table 3.

While the MASC algorithm presented here is, perhaps, only the first step in achieving successful signature extension, it appears clear that the use of a MASC-type algorithm can effect a considerable improvement in the cost-effectiveness of large area crop surveys.

REFERENCES

1. Crane, R.B., W. Richardson, R.H. Hieber, and W.A. Malila, "A Study of Techniques for Processing Multispectral Scanner Data," ERIM Report No. 31650-155-T, September 1973.

2. Hall, F.G., M.E. Bauer, and W.A. Malila, "First Results from the Crop Identification Technology Assessment for Remote Sensing (CITARS)," Proc. of Ninth International Symposium on Remote Sensing of the Environment, April 1974.

3. Henderson, R.G., G.S. Thomas, and R.F. Nalepka, "Methods of Extending Signatures and Training Without Ground Information," ERIM Report No. 109600-16-F, March 1975.

4. Malila, W.A., D.P. Rice, and R.C. Cicone, "Final Report on the CITARS Effort by the Environmental Research Institute of Michigan," ERIM Report No. 109600-12-F, February 1975.

5. Malila, W.A., M.E. Bauer, R.M. Bizzel, F.G.Hall, A. Feiveson, B. Davis, and D.P. Rice, "Results from the Crop Identification Technology Assessment for Remote Sensing (CITARS)," Proc.of Symp. on Machine Processing of Remotely Sensed Data, West Lafayette, Indiana, June 1975.

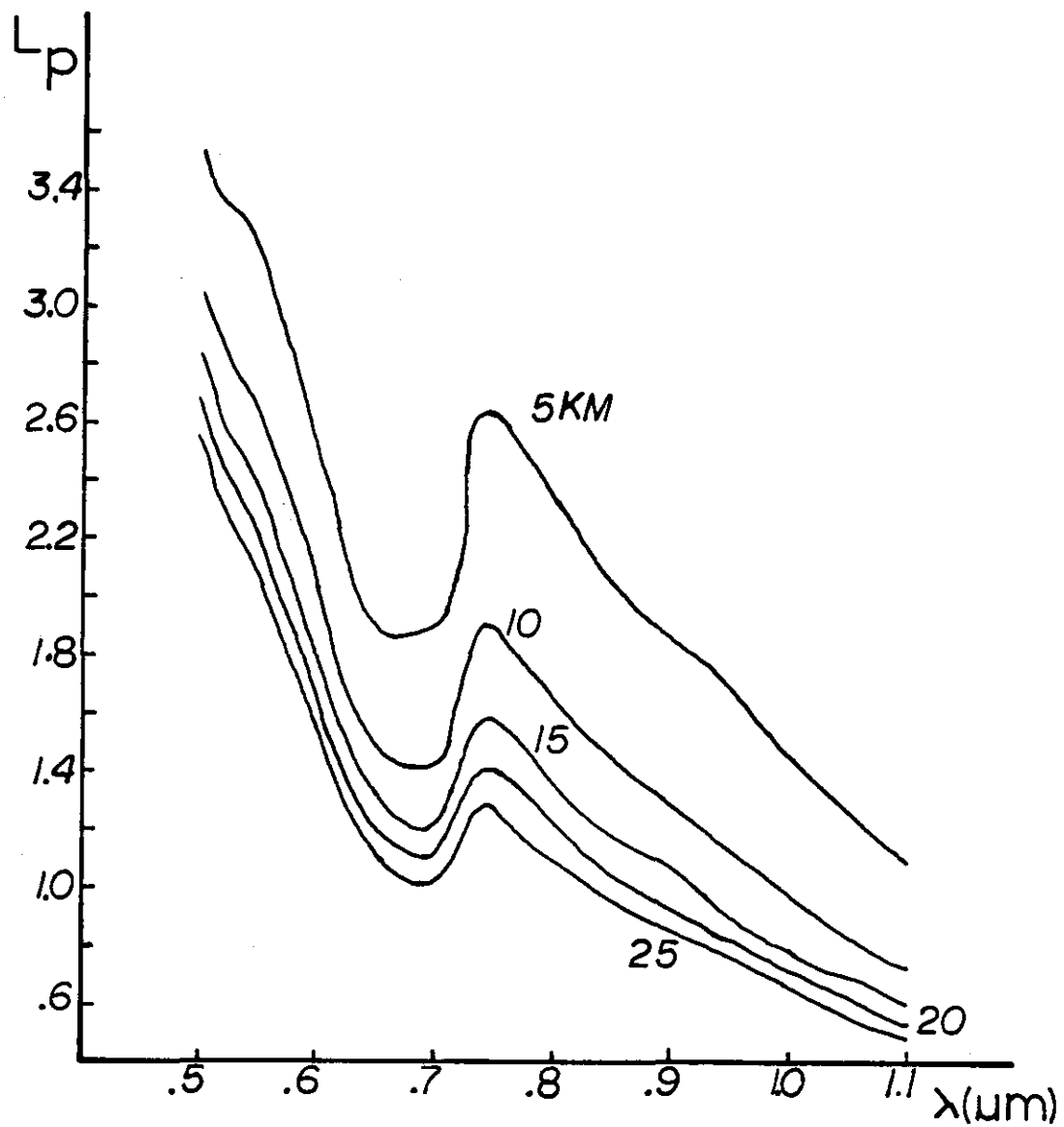6. Turner, R.E., "Radiative Transfer in Real Atmospheres," ERIM Report No. 190100-24-T, December 1973.

Fig. 1. VARIATION OF PATH RADIANCE VS. WAVELENGTH
FOR SEVERAL VISIBILITIES
(Calculated from ERIM Radiative Transfer Model)