




1. Publication Nº <i>INPE-3122-PRE/515</i>	2. Version	3. Date <i>May, 1984</i>	5. Distribution <input type="checkbox"/> Internal <input checked="" type="checkbox"/> External <input type="checkbox"/> Restricted
4. Origin <i>DSC/DIN</i>	Program <i>PROCIM</i>		
6. Key words - selected by the author(s) <i>ENTROPY</i> <i>SPATIAL FEATURES</i> <i>FEATURE SELECTION</i> <i>FILTERING</i> <i>MULTISPECTRAL CLASSIFICATION</i>			
7. U.D.C.: <i>528.711.7:621.376.5</i>			
8. Title <i>EVALUATION OF ENTROPY AND JM-DISTANCE</i> <i>CRITERIONS AS FEATURES SELECTION</i> <i>METHODS USING SPECTRAL AND SPATIAL</i> <i>FEATURES DERIVED FROM LANDSAT IMAGES</i>		10. Nº of pages: <i>09</i>	
		11. Last page: <i>08</i>	
		12. Revised by  <i>Gilberto Câmara Neto</i>	
9. Authorship <i>Luciano V. Dutra</i> <i>F.A. Mitsuo Ii</i> <i>Nelson D. A. Mascarenhas</i>		13. Authorized by  <i>Nelson de Jesus Parada</i> <i>Director General</i>	
Responsible author 			
14. Abstract/Notes <p><i>This research had the purpose of evaluating the performance of entropy and JM-distance feature selection methods, using LANDSAT satellite images. A study area near Ribeirão Preto in São Paulo state was selected, with predominance in sugar cane. Eight features were extracted from the 4 original bands of LANDSAT image, using low-pass and high-pass filtering to obtain spatial features. There were 5 training sites in order to acquire the necessary parameters. Two groups of four channels were selected from 12 channels using JM-distance and entropy criterions. The number of selected channels was defined by physical restrictions of the image analyser and computacional costs. The evaluation was performed by extracting the confusion matrix for training and tests areas, with a maximum likelihood classifier, and by defining performance indexes based on those matrixes for each group of channels. The results showed that in spatial features and supervides classification, the entropy criterion is better ins the sense that allows a more accurate and generalized definition of class signature. On the other hand, JM-distance criterion strongly reduces the misclassification within training areas.</i></p>			
15. Remarks <i>Presented at the International Society for Photogrammetry and Remote Sensing Congress, Rio de Janeiro, June 18-29, 1984.</i>			

EVALUATION OF ENTROPY AND JM-DISTANCE CRITERIONS AS FEATURES
SELECTION METHODS USING SPECTRAL AND SPATIAL FEATURES DERIVED
FROM LANDSAT IMAGES.

Luciano V. Dutra

F.A. Mitsuo Ii

Nelson D.A. Mascarenhas

CNPq - Instituto de Pesquisas Espaciais - INPE

Caixa Postal 515 - São José dos Campos - SP

Brazil

Comission VII W.G. 1

ABSTRACT

This research had the purpose of evaluating the performance of entropy and JM-distance feature selection methods, using LANDSAT satellite images. A study area near Ribeirão Preto in São Paulo state was selected, with predominance in sugar cane. Eight features were extracted from the 4 original bands of LANDSAT image, using low-pass and high-pass filtering to obtain spatial features. There were 5 training sites in order to acquire the necessary parameters. Two groups of four channels were selected from 12 channels using JM-distance and entropy criterions. The number of selected channels was defined by physical restrictions of the image analyzer and computational costs. The evaluation was performed by extracting the confusion matrix for training and tests areas, with a maximum likelihood classifier, and by defining performance indexes based on those matrixes for each group of channels. The results showed that in spatial features and supervised classification, the entropy criterion is better in the sense that allows a more accurate and generalized definition of class signature. On the other hand, JM-distance criterion strongly reduces the misclassification within training areas.

1. INTRODUCTION

One of the main problems in the design of patterns classification systems is the choice of features that should be used to discriminate among the various existing classes.

In the case of pictorial patterns recognition problems, several processes for extraction and selection of features have been developed.

This paper will focus features extration by filtering (spatial features) and feature selection by JM-distance and entropy methods. Several authors have examined different feature selection criterions. Gramenopoulos (1973) employed spatial features derived from filtering the Discrete Fourier Transform over a 32 x 32 window. Ahuja et al. (1977) describe the applications of supervised and nonsupervised methods for image segmentation using gray levels in the neighbourhood of a pixel as features. Schachter et al. (1979) describe some attempts to segment monochromatic images by detecting clusters of certain local features. Logan et al. (1979) synthesized a new channel from LANDSAT channel 5 by calculating the standard deviation in a 3 x 3 window and also utilized that channel for nonsupervised classification in forestry. Dondes and Rosenfeld (1982) extracted features based on gray level fluctuation, measured in the neighbourhood of a pixel and used relaxation techniques to adjust the probabilities for classification. Dutra et al. (1982) describe some experiments, with spatial feature extration in multispectral classification.

This paper reports the use of spectral and new local spatial features in a supervised classification environment. Furthermore, the high dimensionality of the increased feature vector is circumvented by a process of feature selection in order to reduce classification costs. Two methods of feature selection, the JM-distance criterion and the entropy (Chen, 1973) are tested and analysed.

2. SPATIAL FEATURE EXTRACTION

The problem of multispectral image classification in remote sensing has been traditionally approached through spectral features derived from each channel.

However, the task of discrimination is sometimes difficult and the inclusion of spatial attributes can be helpful. Local features can be extracted by filtering, since the spatial frequency content expresses, in some sense, the spatial relationships between pixels.

These filters can be linear or nonlinear and they can enhance different bands of the Fourier spectrum. Figure 1 shows the mask used for linear low-pass filtering. Low frequency components of an image are related with the clustering properties of the classes. This can be explained in terms of the relationship between the value at the origin of the correlation function of a random process and the spectral density function, namely:

$$R(0) = \int_{-\infty}^{+\infty} S(\gamma) D\gamma . \quad (2.1)$$

The use of a low pass filter will tend to decrease the integral on the right side of the Equation 2.1. On the other hand, if zero mean of the random process is assumed, the left side ($R(0)$) is equal to the variance of this process, which is a measure of the scatter of the feature around the mean value.

For extracting roughness information of an image, a heuristic nonlinear filter called "variation" (Schachter et al., 1979) was used by considering a 3 x 3 neighbourhood around a pixel and by labelling the pixels in this neighbourhood by:

a b c
d x e
f g h,

the total variation (T.V.) is the sum of the vertical variation (V.V.) and the horizontal variation (H.V.) i.e.

$$VV = |a-d| + |b-x| + |c-e| + |d-f| + |x-g| + |e-h|, \quad (2.2)$$

$$HV = |a-b| + |d-x| + |f-g| + |b-c| + |x-e| + |g-h|, \quad (2.3)$$

$$TV = VV + HV . \quad (2.4)$$

$$\frac{1}{21} \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

Figure 1 - Low pass filter: average on a 5 x 5 region.

3. FEATURE SELECTION

Images taken by sensors on board remote platforms, as the LANDSAT satellites, are multispectral, each pixel being typified by 5 or more spectral bands.

Although useful for improving class discrimination, the spatial feature extraction processes described in Section 2 can increase the dimensionality of the classification algorithm. This may reduce the computational efficiency and also demand excessive number of samples for training. Therefore, a feature selection process is usually necessary.

In this paper, 2 measures of discrimination are presented and compared. The Jeffreys-Matusita Distance "(JM Distance)" related to the well-known "Bhattacharya Distance" ("B" Distance) -and the not so often used entropy discrimination criterion.

The "B Distance" between two classes w_1 and w_2 described by Gaussian densities is given by Chen (1973):

$$B = \frac{1}{8} (\bar{\mu}_1 - \bar{\mu}_2)^T \frac{(\Sigma_1 + \Sigma_2)}{2} (\bar{\mu}_1 - \bar{\mu}_2) + \frac{1}{2} \left\{ \frac{\frac{1}{2} |\Sigma_1 + \Sigma_2|}{|\Sigma_1|^{\frac{1}{2}} \cdot |\Sigma_2|^{\frac{1}{2}}} \right\}, \quad (3.1)$$

Where μ_i and Σ_i , $i=1,2$ are the mean vector and covariance matrix of class i , respectively. The JM distance is given by Swain et al. (1973).

$$d^2_{JM} = 2(1 - e^{-B}). \quad (3.2)$$

In multiclass problem, the selection is usually made by choosing the set of features that maximizes mean d_{JM} or choosing the set that maximizes the minimum d_{JM} between 2 classes.

For Gaussian patterns the entropy is given by Young and Calvert (1974):

$$H(x) = \frac{1}{2} \ln |\bar{\Sigma}| + \frac{N}{2} \ln 2\pi e, \quad (3.3)$$

where

$|\bar{\Sigma}|$ = determinant of the covariance matrix,

N = number of features.

It is well known that, for Gaussian patterns if one searches for the optimal orthonormal transformation (in the sense of maximizing the entropy for a given dimensionality reduction), the selection is given by the Karhunen-Loeve transform.

In this experiment, however, we shall restrict ourselves to feature selection (i.e. a subset of the nontransformed original features), instead of the more general class of feature extraction methods.

Furthermore, the covariance matrix the Equation 3.3 is the pooled covariance matrix, computed by the average of the covariance matrix of each class weighted by the numbers of points of the training areas. Therefore, the feature selection method will deal with the global distribution of the classes.

Another possibility that has been also explored (Ii et al., 1982) is to assume independence between classes and perform the feature selection by choosing the subset of channels that maximize:

$$S = \sum_{i=1}^M \ln |\Sigma_i| \quad (3.4)$$

where $M = \#$ of classes,

$|\Sigma_i|$ = determinant of the covariance matrix of class i .

4. EXPERIMENTAL RESULTS

The experiments were made with a LANDSAT-C image covering the area of Ribeirão Preto, São Paulo state, Brazil, WRS 236.75, taken on April 1978. Aircraft images from the same area were obtained on June 12th, 1978, at 1:20 000 scale with Kodak Aerochrome 2443 Color IR Film. Ground checks were also made and that allowed a good selection of training and test areas for the classifier.

Six classes were defined 1) sugar cane - 2) new sugar cane - 3) pasture - 4) water - 5) urban development - 6) forest

The main difference between class 1 and class 2 is in the coverage-total in class 1 and partial in class 2 - of the soil by the foliage area.

The number of pixels in the training and test sites is contained in Table 1.

TABLE 1

Number of pixels in training and test Areas

	Number of Pixels	
	Training Area	Test Area
1-Sugar cane	252	108
2-New sugar cane	216	108
3-Pasture	108	72
4-Water	72	36
5-Urban development	72	36
6-Forest	72	36

In the experiment, 12 features were used, according to the following distributions:

- Features 1 to 4 correspond to LANDSAT original channels 4 to 7.
- Features 5 to 8 were obtained by the convolution of channels 4 to 7 with the mask of Figure 1.
- Features 9 to 12 give information about local roughness variations from the original channels. These features were obtained by using the total variation operator defined in Section 2. These channels were further processed with the filter of Figure 1 in order to reduce the effect of noise.

From these 12 features, four were selected by each method, namely maximum global entropy, maximum mean JM-distance and maximum minimum JM-distance between classes.

The 4 channels selected using both JM distance criterious were the same and they are channels 5, 8, 9 and 10. These are two low-pass filtered channels and two high-pass filtered channels.

The four channels selected using the entropy criterious were 4, 10, 11 and 12 first channel being the original band seven, while the other three channels were high-pass filtered channels.

Table 2 presents the average performance (A.P.) defined on the average percentage of correct classification for each site (training areas), weighted by the number of points in the area; the average confusion (A.C.) and the average rejection (A.R.) for training areas.

The L parameters are the rejection threshold on the log likelihood function.

Table 3 presents the same performance indexes for test areas.

TABLE 2

Performance indexes for training areas

	Original Channels		JM Distance		Global Entropy	
	5	6	5	6	5	6
L	5	6	5	6	5	6
A.P.	95.5	95.6	99.5	99.6	93.7	94.8
A.R.	0.4	0.3	0.2	0.3	2.1	0.6
A.C.	4.2	4.2	0.3	0.4	4.2	4.5

TABLE 3

Performance indexes for test sites

	Original Channels		JM Distance		Global Entropy	
	5	6	5	6	5	6
L	5	6	5	6	5	6
A.P.	78.0	80.6	81.1	84.3	91.9	94.9
A.R.	4.8	0.3	13.1	5.8	6.6	1.8
A.C.	17.2	19.2	5.8	9.8	1.5	3.2

The first conclusion that can be drawn is that for any criterion the use of spatial features tends to increase classification accuracy for test sites.

The improvement in performance on the test sites was clearly superior with the use of entropy despite the fact that the global entropy criterion tends to preserve the representation of the distribution of the mixture of classes instead of the discrimination between classes.

Table 4 and 5 below compare this results obtained through the entropy criterion by using Equation 3.3 (entropy of the global distribution) and Equation 3.4 (sum of individual entropies of each distribution which selected channels 4, 9, 11 and 12).

TABLE 4

Performance indexes for training areas

	Global Entropy		Sum of Class Entropy	
	5	6	5	6
L	5	6	5	6
A.P.	93.7	94.8	90.0	94.7
A.R.	2.1	0.6	6.6	2.0
A.C.	4.2	4.5	3.2	3.3

TABLE 5

Performance indexes for test areas

	Global Entropy		Sum of Class Entropy	
	5	6	5	6
L	5	6	5	6
A.P.	91.9	94.9	77.8	91.9
A.R.	6.6	1.8	22.0	6.6
A.C.	1.5	3.3	0.3	1.5

One can notice a marked improvement on the A.P. over test areas by using the global criterion.

In general, the A.R. tended to increase in both training and test areas using any selection criterion, due probably to the fact that these areas include some boundary points in which the variation operator tends to give high output value that are more likely to be rejected.

One should also notice that the JM-distance selection included low-pass filtered channels (with a lower variance) where both entropy criterions did not; This seems to be in accordance with the fact that the JM-distance criterion tends to select features by considering distance between classes and that the entropy criterion, by searching for greater variance (maintaining class representation), tends to select variation operators.

5. CONCLUSIONS

These preliminary results reinforce the importance of the entropy as a feature selection index for remote sensing problems. Further research is needed in this area, for example, by directly calculating the entropy through the histograms of training areas and avoiding the need for the

Gaussian assumption.

REFERENCES

- AHUJA, N.; DAVIS, L.S.; HARALICK, R.M.; PANDA, D.P. *Image segmentation based on local gray level patterns*. Maryland Md, University on Mariland, 1977. (TR551).
- CHEN, C.H. *Statistical pattern recognition*. Rochelle Park, N.J. Spartan Books, 1973.
- DONDES, P.A.; ROSENFELD, A. Pixel classification based on gray level and local "busyness". *IEEE Trans on pattern analysis and machine intelligence*, 4(1):79-84, 1982.
- DUTRA, L.V.; MASCARENHAS, N.D.A. *Some experiments with spatial feature extraction methods in multispectral classification*, Sao Jose dos Campos, INPE, June 1982 (INPE-2443-PRE/144).
- GRAMENOPOULOS, N. Automatic thematic mapping and change detetion of ERTS-1 images. In: EARTH RESOURCES TECHNOLOGY SATELLITE-1 SYMPOSIUM, 3., Washington, DC, 1973. *Proceedings*. Washington, DC, NASA, 1974, V.1B, P. 1845-1875.
- Ii, F.A.M.; DUTRA, L.V.; MENDES, C.L. Comparação entre os métodos de entropia e da distância de Jeffreys-Matusita em problemas de seleção de atributos, *Simpósio Brasileiro de Sensoriamento Remoto*, Brasília, 1982.
- LOGAN, S.L.; STRABLER, A.H. Use of a standard derivation based texture channel for LANDSAT classification of forest strata In: ANNUAL SYMPOSIUM ON MACHINE PROCESSING OF REMOTELY SENSED DATA, West Lafayette, In, 1979. *Proceedings*. New York, IEEE, 1979.
- SCHACHTER, B.J.; DAVIS, L.S.; ROSENFELD, A. Some experiments in image segmentation by clustering of local feature values. *Pattern Recognition* 11(1): 19-28, Jan. 1979.
- SWAIN, P.M.; KING, R.C. *Two effective feature selection criterion for multiespectral remote sensing - West Lafayette*, In, Purdue University, 1973, (LARS information not 402673).
- YOUNG, T.Y.; CALVERT, T.W. *Classification, estimation and pattern recognition*. New York, Elsevier, 1974.