A CLASSIC AND NEURAL PROBABILISTIC APPROACH TO THE DUST STORM DETECTION PROBLEM

P. Rivas-Perea, J. G. Rosiles[*]

The University of Texas El Paso
Dept. of Electrical and Computer Eng
500 W University Ave. El Paso, TX 79968

M. I. Chacon M.[†]

Graduate Studies Department
Chihuahua Institute of Technology
Ave. Tecnologico 2909, Chih. Mex.

ABSTRACT

This paper address the problem of dust storm detection based on multispectral image analysis from a probabilistic point of view. Two classifiers are designed, one based on classic probability theory and other based on a probabilistic computational intelligence approach. The first classifier is designed under the Maximum Likelihood Estimation (MLE) model, and the second with a Probabilistic Neural Network (PNN) model. The data set used in this work consists of MODIS instrument at the NASAs Terra satellite data, generating 75 millions of samples used in the design and validation of the classifiers. Findings indicated that the PNN presents a better classification performance than the MLE classifier. The proposed models are suitable for near real-time applications, and provide with an output at a resolution of 1km, which is an improvement over the methods based on the MODIS AOT product which has a 10km resolution.

Index Terms— Maximum likelihood estimation, Neural network architecture, Image processing, Stochastic approximation, Remote sensing.

1. INTRODUCTION

Nowadays the advances in remote sensing technology make possible the study of different phenomena in order to understand and learn to live with natural phenomena. In our planet there exist natural events that are hazardous for the human being. Most of these events are currently under study. It is well known that the dust storm event air-borne particles (aerosols) are a major cause of several physical, environmental and economical hazards. Air pollution from dust storms is a significant health hazard for people with respiratory diseases and can adversely impact urban areas [1]. Therefore, the study of dust storms events is of big concern in the scientific community.

In spite of the fact that several methods for detecting dust storms exist, there are still open questions in the detection process and in dust storm feature extraction [2]. Current dust storm analysis systems in the state of the art are based in the Moderate Resolution Spectroradiometer (MODIS) Aerosol Optical Thickness (AOT) product from NASA Terra satellite, which typically is available after two days of the observation since its production requires several steps and time. Furthermore, there are no machine learning approaches on image processing techniques that effectively address the dust storm detection problem. Therefore, in this paper we present two models for remote sensing image processing based on probabilistic methods. These methods have a high performance of accuracy and are suitable for dust storm real-time detection applications. These models are based on spectral bands feature extraction on the Moderate Resolution Imaging Spectroradiometer (MODIS) data. The models are based on: Maximum Likelihood Estimation (MLE) model, and the Probabilistic Neural Network (PNN) model. When the models are compared, the neural approach show the best numerical results compared to ground truths from examples found in the literature.

Section 2 of the paper introduces the spectral analysis of dust storms. The MLE and PNN models are explained in Section 3 and 4. Section 5 presents the results of the experimentation followed by a brief discussion of the results and findings. Finally, conclusions are drawn in Section 6.

2. SELECTION OF SPECTRAL BANDS FOR ANALYSIS

The MODIS instrument at the NASA's Terra satellite, is remotely sensing the earth providing information in 36 spectral bands, which allow the analysis of a wide variety of phenomena such as: atmospheric, sea temperature, surface reflectivity, to name only a few. All this data makes possible the analysis of the dust particles suspended in the atmosphere (dust

aerosols). These 36 bands are available in Level 1B of the MODIS file organization. It is intuitive that bands $B1$, $B3$, and $B4$ can be utilized effectively for visual assessment of the dust storms since these bands correspond to the human visual spectra. We can map such bands to an RGB image and produce a true color image given by the following relationship: $R = B1, G = B4, B = B3$. Besides the visual inspection, Hao *et al.*[3] have demonstrated that bands $B20$, $B29$, $B31$, and $B32$ can be utilized effectively for visual enhancement of dust storms. In previous work from Ackerman *et al.*[4] demonstrated that the subtraction of bands $B32$ and $B31$ improves contrast of dust storm images. Based on these findings we selected the following features to design a classification system: $B20$, $B29$, $B31$, $B32$ and $B32 - B31$.

2.1. Dust Storm Data Set

We have selected 31 different events for our experiments. Only 8 out of the 31, were selected for validation. The data was downloaded using NASA's WIST online tool, and the selected events correspond to the south-western US and north-western Mexico area. The selection of which events belong to the modeling or validation set was performed randomly. This data set provides approximately 75 millions feature vectors.

3. PROBABILISTIC MODELING: MULTIVARIATE MAXIMUM LIKELIHOOD CLASSIFIER

Let $f_{C_j}(C_j = c)$ be the probability mass function associated to the j-th class C to occur with a particular value of c. Let $f_{X_{\mathbf{n}}^{(m)}|C_j}\left(X_{\mathbf{n}}^{(m)} = x|C_j = c\right)$ be the conditional probability density function of the $\mathbf{n}$−th pixel of the m-th spectral band to have a value of x given the probability that the j-th class occur with a value of c. This might be referred to as the "a priori" probability. The Maximum Likelihood Classifier (Maximum Likelihood Estimator, MLE), is based on the prior probabilities assuming that the posterior probabilities are unknown. The MLE is an accepted measure for classification and analysis of remotely sensed data [5]. Therefore we modeled an MLE classifier specialized on classification of dust and background. Our model uses four classes: $C0$ for land/sea background, SM for smoke, $BLDU$ for blowing dust, and DS for dust storm. The probability mass functions are expressed as density functions for convenience and will be estimated from the collected samples.

The MLE can be derived from Bayes theorem. Since we are interested on finding the maximum likelihood between the observed data and the prior probability for all classes, we can state a decision rule as follows

$$x \in C_j \text{ if, } \quad f_{X_{\mathbf{n}}^{(m)}|C_j} f_{C_j} > f_{X_{\mathbf{n}}^{(m)}|C_i} f_{C_i}, \quad \forall j \neq i \quad (1)$$

which appear already simplified by simply removing the common factor $f_{X_{\mathbf{n}}^{(m)}}\left(X_{\mathbf{n}}^{(m)} = x\right)$. Then, if we assume that the

prior probability is normally distributed we can rewrite the above terms by defining

$$\psi_{k,\mathbf{n}}^{(m)}(x) = f_{X_{\mathbf{n}}^{(m)}|C_k} f_{C_k} \quad (2)$$

that allows restating the decision rule as follows

$$x \in C_j \quad \text{if} \quad \psi_{j,\mathbf{n}}^{(m)}(x) > \psi_{i,\mathbf{n}}^{(m)}(x) \quad \forall j \neq i \quad (3)$$

which is more convenient to handle. In MLE, the functions $\psi_{k,\mathbf{n}}^{(m)}(x)$ are commonly referred to as *discriminant functions*.

The discriminant functions $\psi_{k,\mathbf{n}}^{(m)}(x)$ can be reduced because of the Gaussianity assumption by removing the factor $-\frac{d}{2}\ln(2\pi)$ which adds no discriminant information to the classification [6], and also since the uncertainty is high for the true PMF $f_{C_k}(C_k = c)$, it is removed from (2), as well as the common factor $\frac{1}{2}$, leading to a commonly used simpler discriminant function

$$\psi_{k,\mathbf{n}}^{(m)}(x) = -\det\left(\mathbf{\Sigma}_{X_{\mathbf{n}}^{(m)}|C_k}\right) - \dots$$
$$\left(x - \mu_{X_{\mathbf{n}}^{(m)}|C_k}\right)^T \mathbf{\Sigma}_{X_{\mathbf{n}}^{(m)}|C_k}^{-1} \left(x - \mu_{X_{\mathbf{n}}^{(m)}|C_k}\right) \quad (4)$$

where $\det(\cdot)$ is the determinant function, $\mathbf{\Sigma}_{X_{\mathbf{n}}^{(m)}|C_k}$ is the covariance matrix of the prior PDF, $\mu_{X_{\mathbf{n}}^{(m)}|C_k}$ denotes the mean vector, $(\cdot)^T$ denotes the transpose operation, and $\mathbf{\Sigma}_{X_{\mathbf{n}}^{(m)}|C_k}^{-1}$ is the inverse of the covariance matrix $\mathbf{\Sigma}_{X_{\mathbf{n}}^{(m)}|C_k}$.

3.1. Features for Recovered Radiance Analysis

We use the term "Recovered" to refer 16 bit MODIS data that is recovered to its original scale (W/m^2/μm/sr). The recovery process is given by

$$L_{\mathbf{n}}^{(m)} = \kappa^{(m)}(\iota_{\mathbf{n}}^{(m)} - \eta^{(m)}) \quad (5)$$

where $L_{\mathbf{n}}^{(m)}$ denotes the recovered radiances, $\kappa^{(m)}$ are the radiance scales, $\eta^{(m)}$ are the radiance offsets, and $\iota_{\mathbf{n}}^{(m)}$ are the scaled intensities (raw data). A feature vector $F \in \Re^{n \times 5}$ is constructed from the recovered radiances as follows

$$F = \left[L^{B20}, L^{B29}, L^{B31}, L^{B32}, \left(L^{B32} - L^{B31}\right)\right]. \quad (6)$$

The feature $L^{B32} - L^{B31}$ emphasizes that the subtraction of $B31$ from $B32$ is performed after the recovery process.

3.2. Estimation of the Conditional PDF $f_{X_{\mathbf{n}}^{(m)}|C_j}$

Considering Gaussianity, the goal is to estimate the following parameters: the vector of expected values $\widehat{\mu}_{F|C_j}$ and the covariance matrix $\widehat{\mathbf{\Sigma}}_{F|C_j}$. In order to perform the estimation, we performed segmentation based on images that have been already published on remote sensing research journal, and from the images contained in these papers, we have identified the dust storm region and manually segmented the sample images containing dust storms, blowing dust, and smoke,

such that we can utilize the region of the segmented image as a mask. This mask is associated with a particular class C_j. Then having all the masks and the associated classes, we extract and store the subset of pixels associated to the $j-$th class. Finally, we compute the sample mean $\widehat{\mu}_{F|C_j}$ and the covariance matrix $\widehat{\Sigma}_{F|C_j}$ = which is clearly not ill-posed since we have a large number of data samples available for modeling as in most remote sensing applications.

4. NEURO-PROBABILISTIC MODELING: THE PROBABILISTIC NEURAL NETWORK

The Probabilistic Neural Network (PNN) is a semi-supervised neural network widely used in pattern recognition applications. One of the main advantages is that it does not require training. Indeed, the PNN is inspired in the Bayesian classification. The basic operation of the PNN is to estimate the PDF's of the features assuming Gaussian distributions. The general architecture of the PNN is composed of four layers. The first layer is an input layer receiving the features $F \in \Re^n$. The second layer contains exponential functions $\varphi(\cdot)$ in each nodes, and the number of nodes correspond to the k number of samples available for training for the $j-$th class. This nodes are called pattern units ν_{jk}^F, and are fully connected to the input nodes. The output of the pattern layer is denoted by

$$\varphi_{jk}(F) = \frac{1}{(2\pi)^{\frac{d}{2}} \sigma^d} e^{-\frac{1}{2\sigma^2} (F - \nu_{jk}^F)^T (F - \nu_{jk}^F)}. \quad (7)$$

The third layer contains summation units to complete the probability estimation. There are as many summation units as classes. The $j-$th summation unit denoted as ϖ_j, receives input only from those pattern units belonging to the $j-$th class. This layer computes the likelihood of F being classified into C_j, by averaging and summarizing the output of all neurons that belong to the same class as

$$\Omega_j (\varphi_{jk}(F)) = \frac{1}{(2\pi)^{\frac{d}{2}} \sigma^d} \frac{1}{N_j} \times \dots$$
$$\sum_{i=1}^{N_j} e^{-\frac{1}{2\sigma^2} (\varphi_{ik}(F) - \varpi_i)^T (\varphi_{ik}(F) - \varpi_i)}. \quad (8)$$

The last layer is the decision layer. It classifies the pattern of F according to the Bayesian decision rule given by

$$F \in C_j \text{ if, } C_j (\Omega_j (\varphi_{jk}(F))) = \max_{1 \le i \le j} \Omega_i (\varphi_{ik}(F)). \quad (9)$$

Thus, the maximum of the summation node outputs can be expressed as a function of $C_j (\cdot)$ characterizing the output of this layer.

4.1. Features and Estimation of the Spread Parameter σ

The input to the PNN model is F in (6). The parameter σ_F is estimated by the method proposed by Srinivasan et al.[7],

which requires a phase of pre-normalization of the data consisting on subtracting the mean μ_F from the training feature vector F, and dividing it by its standard deviation σ_F. The value of σ_F is defined as the absolute difference between the two smallest normalized variances. This completes the model of the PNN since there is no training phase involved.

4.2. The problem of the large sample size in modeling the PNN

In order to avoid the overwhelming processing of millions of training samples, we decided to limit the number of samples selected for training in the case of PNN. We based our reduction method on the criteria that establishes that the number of samples required for training the networks must be at least 3 times the number of bands used as features [8]. Therefore, in the PNN design, we decided to use at least 3 times the size of the feature vector F. The selection of which feature vectors to keep will not be addressed in this paper.

5. RESULTS AND DISCUSSION

The selected performance metrics are considering all samples used in the different methods. These metrics are:

$$\text{Precision} = \frac{\sum TP}{\sum TP + FP}, \quad (10)$$

$$\text{Accuracy} = \frac{\sum TP + TN}{\sum TP + FN + FP + TN}, \quad (11)$$

as well as the area under the receiver operating characteristics (ROC) curve (AUC). AUC is widely used metric because its superiority in reflecting the true performance of a classification system. The numerical results were concentrated and averaged to produce Table 1, showing that the neuro-probabilistic approach is better than the MLE. Table 1 also include the processing time per feature vector in milliseconds. Visual results of the classification are illustrated in Figure 1. Figure 1 a) is a true color image obtained with the traditional Mercator approach. Figure 1 b) shows the probability of presence of dust storm. Figure 1 c) is a segmentation image, showing the actual mapping. DS is mapped to red, $BLDU$ to green, SM is blue, and $C0$ is black.

6. CONCLUSION

The problem of dust storm detection has been addressed in this paper. We have modeled probabilistic approaches for dust storm detection and classification. These models are specialized on measuring the probability of the presence of dust storm data given MODIS Level 1B data. Novel techniques in Machine Learning were utilized to design a neural architecture to model dust storms. To the best of the knowledge of the authors, the presented models are the first in its

kind that can actually perform classification of dust storms pixels based on soft computing methods. We compared the probabilistic models such as the Maximum Likelihood, MLE, and the Probabilistic Neural Network, PNN. The PNN show a strong ability in inferring the relationship between spectral bands to classify dust, and discriminate from other classes, such as clouds, smoke, etc. Moreover, the proposed probabilistic models are suitable for near real-time applications, such as direct broadcast, rapid response analysis, emergency alerts, etc. The work reported in this document is suitable for the study of the dust storm problem since the algorithms can show the dust presence to a resolution of 1km. This represents an improvement over the methods based on the Aerosol Optical Thickness index (AOT) which lack of resolution and its results are generated after two days of the satellite pass.

7. REFERENCES

[1] Rivera Rivera, N.I., Gebhart, K.A., Gill, T.E., Hand, J.L., Novlan, D.J., and Fitzgerald, R.M., 2009. Analysis of air transport patterns bringing dust storms to El Paso, Texas. Symposium on Urban High Impact Weather, AMS Annual Meeting, January 2009, Phoenix., JP2.6.

[2] Khazenie, N.; Lee, T.F., "Identification Of Aerosol Features Such As Smoke And Dust, In NOAA-AVHRR Data Using Spatial Textures," Geosc. and Rem. Sens. Symp., pp.726-730, 26-29 May 1992.

[3] Xianjun Hao, John J. Qu, 2007, Saharan dust storm detection using MODIS thermal infrared bands. Journal of Applied Remote Sensing, Vol. 1, pp. 013510.

[4] Ackerman, S.A., Strabala, K.I., Menzel, W.P., Frey, R., Moeller, C., Gumley, L., Baum, B.A., Seeman, S.W., and Zhang, H.. Discriminating clear-sky from cloud with MODIS algorithm theoretical basis document (MOD35). ATBD-MOD-06, version 4.0, NASA Goddard Space Flight Center.

[5] Aksoy, S.; Koperski, K.; Tusk, C.; Marchisio, G.; Tilton, J.C., "Learning bayesian classifiers for scene classification with a visual grammar," Geosc. and Rem. Sens., IEEE Trans. on, vol.43, no.3, pp. 581-589, 2005.

[6] John A. Richards and Xiuping Jia, Remote Sensing Digital Image Analysis, Springer, 2006.

[7] Ramakrishnan, S. and Selvan, S. Image texture classification using wavelet based curve fitting and probabilistic neural network. Int. J. Imaging Syst. Technol. 17, 4 pp. 266-275. 2007.

[8] Kanellopoulos, G. G. Wilkinson, J. Austin, Neurocomputation in Remote Sensing Data. Analysis, Springer-Verlag, 1997.
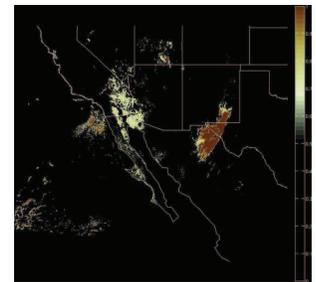
Table 1. Classifiers Performance.

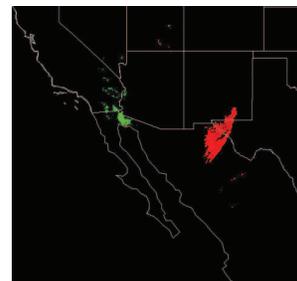|	Precision	Accuracy	AUC	P. Time
MLE	0.5255	0.6779	0.4884	0.0141
PNN	0.7664	0.8412	0.6293	0.2393
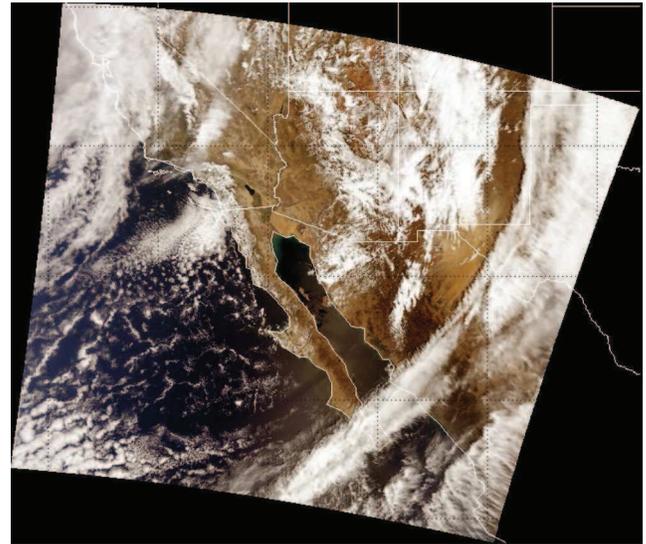
(a) True color image R=B1, G=B4, and B=B3
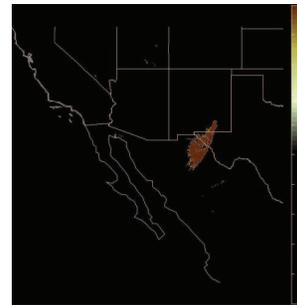
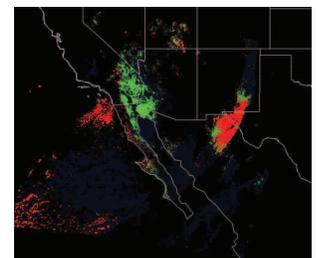(b) Dust probability MLE

(c) Dust probability PNN

(d) Segmentation MLE

(e) Segmentation PNN

Fig. 1. Dust storm event on April 6th 2001.