

**ENVIRONMENTALLY-ORIENTED PROCESSING OF
MULTI-SPECTRAL SATELLITE IMAGES:
NEW CHALLENGES FOR BAYESIAN METHODS**

S.A. STARKS AND V. KREINOVICH
*NASA Pan American Center
for Earth and Environmental Studies
University of Texas at El Paso
El Paso, TX 79968, USA[†]*

Abstract. Remotely sensed images from new generation satellites present an opportunity for scientists to investigate problems in environmental and earth science which have been previously intractable. The magnitude of data that will arise from these hyperspectral instruments create the need for innovative techniques to accomplish data reduction. This paper presents an algorithm which shows promise as a tool for reducing the dimensionality of data resulting from remote sensing. The optimality criteria for the algorithm is the Bayes Risk in the reduced dimension space.

Key words: satellite imaging, multi-spectral satellite data, environmental applications, Bayes risk

1. Introduction

Recently, the University of Texas at El Paso, through a grant from the National Aeronautics and Space Administration (NASA), created the Pan American Center for Earth and Environmental Studies (PACES). As part of its goals, PACES seeks to establish a repository of remotely sensed and ground-based data to enable scientific studies of the geologic, ecologic and environmental processes taking place in the area of the Southwestern United States and Northern Mexico [1].

Current remote sensing imagery result from the processing of data obtained from sensors mounted on satellites with the capability of taking measurements in a limited number of spectral bands. A well known and widely used source of remotely sensed imagery is the Landsat series of satellite launched by NASA [2]. The maximum number of bands associated with the Landsat series of platforms

[†]Emails: {sstarks,vladik}@utep.edu

is seven. As a result, for each point on the earth which we wish to study, we are limited to seven measurements arising from different regions of the electromagnetic spectrum. This limitation on the number of measurements often prevents the study of interesting phenomena because the information which we can obtain via remote sensing is not sufficient to discriminate different earth features. To improve the ability to gather information about our planet, NASA is in the process launching new generation satellites under its Mission to Planet Earth strategic enterprise [3]. Many of these systems will be based upon hyperspectral sensing which will enable the collection of data in the order of several hundred spectral bands [4].

The drastic increase in the amount of data that will be collected make it extremely important to look for new data compression and feature extraction techniques. Currently, a mix of techniques are employed in this area ranging from the *heuristic* to the highly mathematical. Some are based upon elements drawn from soft computing techniques such as neural nets. In this work, we present an algorithm which is based upon the minimization of the Bayes risk of misclassification which we intend to investigate as a means for reducing the dimensionality of hyperspectral imagery.

2. Mathematical Background

In the study of statistical pattern recognition, much attention has been focused on the field of dimensionality reduction. Numerous algorithms have been developed which are optimal with respect to a variety of criteria. Some criteria include measures of pattern class separability such as the Bhattacharyya distance and the average divergence. Other techniques have their bases in the field of discriminant analysis. Still other methods are based upon the Karhunen-Loeve expansion and factor analysis. Since the ultimate test of any practical pattern recognition system is its ability to perform accurate classification in the reduced-dimension or feature space, it seems natural that a desirable figure of merit for the dimensionality reduction algorithm is the Bayes risk of misclassification in the feature space. The mathematic ground work for this work was laid in a paper by de Figueiredo [5].

In this report, an algorithm for determining an optimal linear transformation \tilde{A} from a real n -dimensional Euclidean measure space E^n , in which the raw data to be classified into M ($M \geq 2$) Gaussian pattern classes appear, to a feature space E^m of pre-described dimension $m < n$, in which classification is to be performed. We denote the latter space by $E^m(\tilde{A})$ to emphasize that the probability measures in E^m depend upon the transformation \tilde{A} .

The case in which the dimensionality of the feature space m is unity warrants special treatment and is described in [6]. In the present paper, we describe an approach for the case in which $m \geq 2$.

Let H^1, \dots, H^M denote the M pattern classes under consideration. We assume that the class conditional probability density functions $f_X(x | H^j)$, $j = 1, \dots, M$ on E^n to be of the form

$$f_X(x | H^j) = (2\pi)^{-n/2} |\tilde{R}^j|^{-1/2} \exp \left[-\frac{1}{2} (x - \bar{x}^j)^T (\tilde{R}^j)^{-1} (x - \bar{x}^j) \right] \quad (1)$$

for $j = 1, \dots, M$. Here, $x = \text{col}(x_1, \dots, x_n)$, and the $(n \times 1)$ vector \bar{x}^j and the $(n \times n)$ symmetric positive definite matrix \tilde{R}^j are the mean and covariance of H^j in E^n . The superscript T denotes the transpose operator and $|\tilde{R}^j|$ denotes the determinant of \tilde{R}^j .

Upon application of the linear dimensionality reduction transformation $A : E^n \rightarrow E^m$, the class conditional probability density functions given in (1) are converted to the corresponding ones in E^m :

$$f_Y(y | H^j) = (2\pi)^{-n/2} |R^j|^{-1/2} \exp \left[-\frac{1}{2} (y - \bar{y}^j)^T (R^j)^{-1} (y - \bar{y}^j) \right], \quad (2)$$

where

$$y = Ax \in E^m(A), \quad (3)$$

$$\bar{y}^j = A\bar{x}^j, \quad (4)$$

and

$$R^j = A\tilde{R}^j A^T. \quad (5)$$

We assume that the classification strategy in the feature space $E^m(A)$ is Bayes.

In what follows, we present an algorithm which, for a given dimensionality $m \geq 2$ of the feature space $E^m(A)$, obtains the optimal transformation \hat{A} by minimizing the Bayes risk of misclassification $R_m(A)$ in $E^m(A)$ over the set of all linear transformations $A : E^n \rightarrow E^m$.

3. Description of the Algorithm

The approach used in the formulation of the algorithm under discussion is based upon minimizing the Bayes risk in the feature space under a constraint placed upon the linear transformation matrix A . We may write the Bayes risk in as

$$R_m(A) = \sum_{i=1}^M \sum_{\substack{j=1 \\ j \neq i}}^M P_j C_{ij} \int_{\Omega_i(A)} f_Y(y | H^j, A) dy, \quad (6)$$

where the probability density functions f_Y are expressed as in (2), $\Omega_i(A)$ is the Bayesian decision region in $E^m(A)$ for class H^i , and the non-negative numbers C_{ij} constitute the elements of the Bayes cost matrix.

If

$$C_{ij} = 1 - \delta_{ij}, \quad (7)$$

where δ_{ij} is the Kronecker delta, (6) represents the probability of misclassification in $E^m(A)$.

Without loss of generality, we require that the norm of the transformation matrix A remain constant for the duration of the optimization procedure. With this constraint, we construct from (6), under (7), an expression for the Lagrangian

$$R_m(A) = \sum_{i=1}^M \sum_{\substack{j=1 \\ j \neq i}}^M P_j \int_{\Omega_i(A)} f_Y(y | H^j, A) dy + \lambda \left(\frac{1}{2} \text{trace}(AA^T) - \alpha \right), \quad (8)$$

where λ is the Lagrange multiplier for the problem, and α is a positive constant usually set equal to unity.

The algorithm determines the optimal dimensionality reduction transformation \widehat{A} , by calculating the extremum of (8) with respect to A . This is done by the iterative procedure presented in the following text. At each step of the approach an estimate of the gradient of (8) with respect to the elements of the transformation matrix A is needed. We may express this gradient as

$$\nabla_A Q_m(A, \lambda) = \sum_{i=1}^M \sum_{\substack{j=1 \\ j \neq i}}^M P_j \left\{ [((R^j)^{-1} D^{ij} - E^{ij} I) \cdot (R^j)^{-1} A \widetilde{R}^j + (R^j)^{-1} F^{ij} (\bar{x}^j)^T] \right\} + \lambda A \quad (9),$$

where

$$D^{ij} = \int_{\Omega_i(A)} (y - \bar{y}^j)(y - \bar{y}^j)^T f_Y(y | H^j, A) dy, \quad (10)$$

$$E^{ij} = \int_{\Omega_i(A)} f_Y(y | H^j, A) dy, \quad (11)$$

$$F^{ij} = \int_{\Omega_i(A)} (y - \bar{y}^j) f_Y(y | H^j, A) dy, \quad (12)$$

and I is the $(m \times m)$ identity matrix.

The derivation of the above expression (9) is lengthy and is presented elsewhere (see, e.g., [5]).

For a given dimensionality $m \geq 2$ for the feature space, our approach replaces the integrals expressed in (10)–(12) by their estimates, which are provided below:

$$D^{ij} = \frac{1}{N_j} \sum_{y_k^j \in \Omega_i(A) \cap S^j} (y - \bar{y}^j)(y - \bar{y}^j)^T, \quad (13)$$

$$E^{ij} = \frac{1}{N_j} \sum_{y_k^j \in \Omega_i(A) \cap S^j} 1, \quad (14)$$

$$F^{ij} = \frac{1}{N_j} \sum_{y_k^j \in \Omega_i(A) \cap S^j} (y - \bar{y}^j), \quad (15)$$

where

$$S^j = \{y_k^j \in E^m(A) \mid k = 1, \dots, N^j\} \quad (16)$$

is a set of N^j labeled samples drawn from the class H^j .

4. Implementation of the Algorithm

The algorithm for determining the optimal transformation matrix \widehat{A} is based upon an iterative procedure which makes use of the expressions for the gradient of

the criterion expressed in (9). The algorithm makes use of a set of labeled n -dimensional samples drawn from a training set to calculate the value of the criterion function as well as that of its gradient for a particular choice of the transformation matrix. First the labeled samples are transformed into the m -dimensional feature space using an initial guess for the transformation matrix A_0 . The samples are then classified in the reduced-dimension space $E^m(A_0)$, according to the Bayes decision scheme. The value for the probability of error is then estimated using the formula

$$\tilde{P}_E(A_0) = \sum_{i=1}^M \sum_{\substack{j=1 \\ j \neq i}}^M P_j \frac{\beta_{ij}}{\delta_j} \quad (17)$$

where β_{ij} is equal to the number of samples from the training set representative of class H^j which are classified in $E^m(A_0)$ as being from class H^i , and δ_j represents the total number of samples taken from the set representative of class H^j .

The value for the criterion function is then calculated by replacing the integral in (8) by $\tilde{P}_E(A_0)$. The value for the gradient of (8) is also calculated using the samples from the training set and the formulas (9), (13), (14), and (15).

An unconstrained multivariate optimization procedure such as that developed by Fletcher and Powell is then invoked to minimize the criterion function by adjusting the elements of the matrix A_0 based upon the value of the gradient. This process is repeated until convergence is obtained.

The basic algorithm for computing the optimal transformation \hat{A} and its corresponding Lagrange multiplier $\hat{\lambda}$ which minimize (8) is presented as follows:

Basic Optimal Linear Feature Extraction Algorithm:

- (a) Enter.
- (b) Obtain class statistics through training.
- (c) Obtain initial A_0 and λ_0 ; set $k \leftarrow 0$.
- (d) Optimize $Q_1(A, \lambda_k)$ by varying A to form A_{k+1} .
- (e) Convergence test.
 - If convergence is obtained, we return $\hat{A} \leftarrow A_{k+1}$ and $\hat{\lambda} \leftarrow \lambda_k$.
 - If convergence is not obtained, we go to block (f).
- (f) Update λ_k to form λ_{k+1} .
- (g) Set $k \leftarrow k + 1$, and go to block (d).

It is important to note that at each iterative step, the value of the Lagrange multiplier is updated by the formula

$$\lambda_{k+1} = (\langle A_{k+1}, A_{k+1} \rangle)^{-1} \left\{ \frac{1}{2} (\|A_{k+1}\|^2 - 1) - \langle A_{k+1}, \nabla_A f(A_{k+1}) \rangle \right\}, \quad (18)$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product in E^n , $\|\cdot\|$ represents the Frobenius norm in E^n , and

$$f(A) = Q_m(A, \lambda) - \lambda \left(\frac{1}{2} \text{trace}(AA^T) - \alpha \right). \quad (19)$$

Also, it should be noted that in block (d) of the above algorithm, a single iteration of the Fletcher-Powell minimization routine is performed by leaving the $(m \times n)$ components of the transformation matrix A_k free to vary and holding λ_k constant. A convergence test may be applied to check whether the optimal A_k and λ_k have been found. The convergence test bases its result on the change in A_k from one step to the next and on the norm of the gradient.

5. Summary and Future Work

An iterative algorithm has been presented for the development of a linear dimensionality reduction transformation for the case that the pattern classes are multi-variate Gaussian and the decision strategy is Bayes. Current work focuses upon other parametric forms of transformation where the class-conditional probability density functions are non-Gaussian. Application of the current approach to hyperspectral remotely sensed imagery is under development.

Acknowledgment

This work was supported in part by NASA under cooperative agreement NCCW-0089. It was also partially supported by NSF under grants No. DUE-9750858 and EEC-9322370, and by the Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant number F49620-95-1-0518.

References

1. S.A. Starks, G.R. Keller, and D.E. Cooke, "Earth and Environmental Remote Sensing at PACES," *Geocarto, Int'l.*, 1997, Vol. 12, No. 3 (Sept. 1997).
2. T.M. Lillesand and R.W. Kiefer, *Remote Sensing and Image Interpretation*, J. Wiley, New York, 1994.
3. G. Asrar and R. Greenstone, eds., *1995 MTPE EOS Reference Handbook*, NASA Goddard Space Flight Center, Greenbelt, MD, 1995.
4. Schott, *Remote Sensing: The Image Chain Approach*, Oxford University Press, New York, 1997.
5. R.J.R. De Figueiredo, "Optimal Linear and Nonlinear Feature Extraction from Several Gaussian Pattern Class," *Proc. of Second Joint Int'l Conf. On Pattern Recognition*, Copenhagen, Denmark, August 1974.
6. S.A. Starks, R.J.R. de Figueiredo, and D.L. Van Rooy, "An Algorithm for Optimal Single Linear Feature Extraction from Several Gaussian Pattern Classes," *Int'l. J. of Computer and Info. Sciences*, 1977, Vol. 6, No. 1, pp. 41-54.