Analytical design of multispectral sensors

Hyperspectral session - Landgrebe

Miguel A. Veganzones

Grupo de Inteligencia Computacional
Universidad del País Vasco

2012-03-09

Outline

Outline

Paper

180 IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. GE-18, NO. 2, APRIL 1980

Analytical Design of Multispectral Sensors

DANIEL J. WIERSMA, MEMBER, IEEE, AND DAVID A. LANDGREBE, FELLOW, IEEE

Abstract—An analytical procedure for the design of the spectral channels for multispectral remote sensor systems is defined. An optimal design based on the criterion of minimum mean-square representation error using the Karhunen–Loeve expansion was developed to represent the spectral response functions from a stratum based upon a stochastic process scene model. From the overall pattern recognition system perspective the effect of the representation accuracy on a typical performance criterion, the probability of correct classification, is investigated. The optimum sensor design provides a standard against which practical (suboptimum) operational sensors can be compared. An example design is provided and its performance is illustrated.

Although the analytical technique was developed primarily for the purpose of sensor design it was found that the procedure has potential for making important contributions to scene understanding. It was concluded that spectral channels which have narrow bandwidths relative to current sensor systems may be necessary to provide adequate spectral representation and improved classification performance.

Motivation

- Define an analytical procedure for the design of the spectral chan- nels for multispectral remote sensor systems.
  - An optimal design based on the criterion of minimum mean-square representation error using the Karhunen-Loeve expansion.
  - Study the effect of the representation accuracy on the probability of correct classification.
- Goals:
  - Primary: the optimum sensor design provides a standard against which practical (suboptimum) operational sensors can be compared.
  - Secondary: the procedure also has potential for making important contributions to scene understanding.
- Conclusion: spectral channels which have narrow bandwidths relative to current sensor systems may be necessary to provide adequate spectral representation and improved classification

Outline

Issues

In order to achieve an optimal design of a set of spectral features one must have suitable analytical representations for 1) the spectral response of the scene, 2) the sensor system, 3) the processor system, and 4) one must have a suitable, analytically expressible optimality criterion. Further, we note the following factors which influence the creation of a spectral feature design procedure.

Spectral response of the scene

1) The scene is very complex in the fashion in which it reflects and emits optical radiation. Mathematical models which predict the scene radiant exitance at least to the level of accuracy and precision needed for our problem, do not yet exist. As a result an empirical scene model must be used.

Sensor system

2) Because satellite-borne sensor systems are very expensive they cannot usually be designed specifically for a certain use or user. Rather they must be optimized with regard to a large number of scenes and uses (Fig. 2). The feature space which the sensor defines must be adequately detailed, for example, such that in early season when agricultural crop canopies have achieved only 10–15-percent cover, both the crop species mapping user and the soils mapper can be served. This fact is important in the choice of optimality criterion, as will be seen shortly.

Sensor system

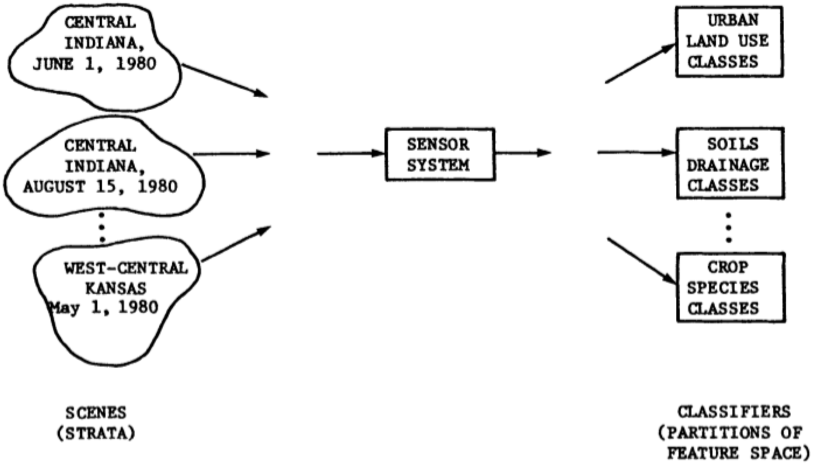


Fig. 2. A sensor system must be designed to perform satisfactorily for

Processor system

3) It is highly desirable that the spectral features be designed in such a way that they are maximally efficient in the sense that a feature set of any given size contain the maximum amount of useful information possible so that any given analysis can proceed with the smallest number of features possible. There are at least three reasons for this: feature efficiency in this sense tends to decrease the amount of processor computation required, it tends to decrease the processor complexity required, and it tends to reduce the amount of training sample data needed.

Optimality criterion

4) There are a number of constraints on the design of a sensor, generally of a practical character, which cannot reasonably be expressed analytically. Examples are those resulting from optical design considerations, sensor material sensitivity curves, cost factors, spacecraft size and weight considerations, etc., and especially from the interrelationships of these types of factors. To mitigate this circumstance we will use the scheme depicted in Fig. 3. That is, we will determine optimal sensor characteristics using entirely analytical means without regard for their physical realizability. These characteristics will then serve as a guide by which to determine nearly equivalent but physically realizable characteristics which perform nearly as well.

Optimality criterion

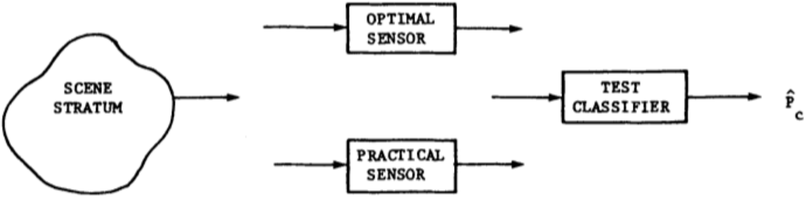


Fig. 3. Sensor system design procedure.

Outline

Spectral response function

Let us begin by considering the information bearing aspects of the spectral response function $x(\lambda)$ [8]. This response function (e.g., from a single pixel) is proportional to the electromagnetic energy received by the sensor as a function of wavelength λ (Fig. 4). Many factors determine the spectral response function for a given observation. The irradiance of the sun, the conditions of the atmosphere, and the reflectance of the surface features all have important effects on the response. Since a deterministic relationship between the response function and the many factors affecting it would be very complex, the set of functions which are observed in practice are best modeled as a stochastic process.

Spectral response function

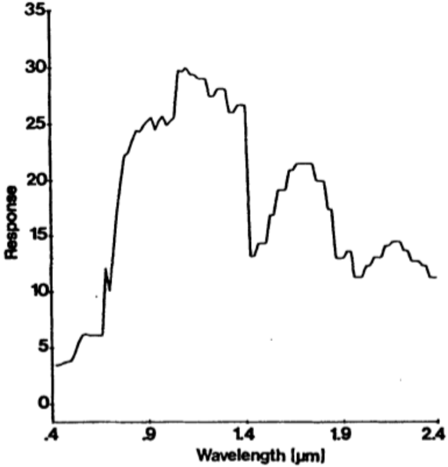


Fig. 4. Spectral response function for Mature Wheat collected on

Stratification

The ensemble of the stochastic process [13] will be defined in terms of the stratification necessary to apply pattern recognition methods to the earth observational problem. A stratum S is defined as the largest contiguous area which can be classified to an acceptable level of performance with a single training of the classifier. It is noted that the sensor must be designed to operate satisfactorily over a large number of such strata, which vary greatly with time, location and application. The collection of all possible strata which a sensor may observe is denoted by S_0. Since the set S_0 is quite large, it is necessary to select a smaller subset which is representative in a statistical sense in order to perform the design.

Stochastic process

The random experiment for the stochastic process consists of the observation of a point in a stratum *S*. Each point in the stratum is mapped into a spectral response function (Fig. 5). The collection of all response functions from a stratum defines an ensemble. The ensemble plus the corresponding probability measure defines the stochastic process [13]. It is appropriate to assume a Gaussian probability measure for this process [3].

Stochastic process

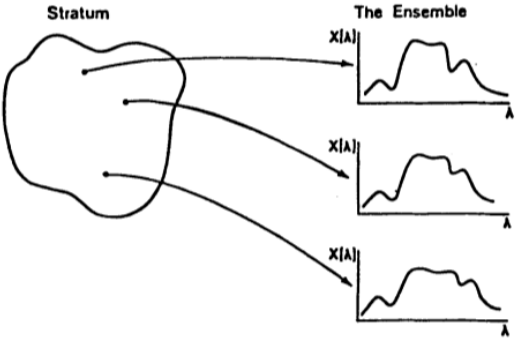


Fig. 5. Realization of a stratum as the ensemble of spectral sample functions.

Outline

Sensor mathematical model

Next we choose a mathematical model for the sensor to represent the spectral response function for each observation. Let the sensor be represented by a set of N filter functions or basis functions $\{\phi_i(\lambda)\}$ such that the output of each filter is given by (Fig. 6)

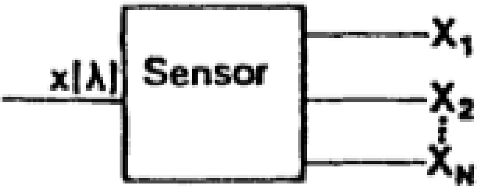
$$x_i = \int_\Lambda x(\lambda)\,\phi_i(\lambda)\,d\lambda. \qquad (1)$$

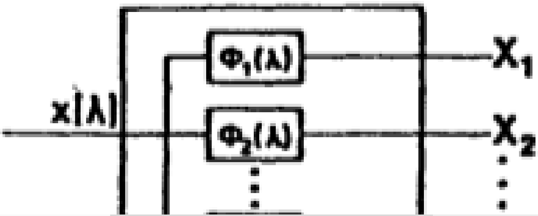
The output of the sensor model is a sequence, $\{x_1, x_2, \cdots, x_N\} = X$, which represents the spectral response by the approximation

$$x(\lambda) \approx x_1\phi_1(\lambda) + x_2\phi_2(\lambda) + \cdots + x_N\phi_N(\lambda)$$

N

Sensor mathematical model





Example

A simple illustration of the concept is given in Fig. 7. However, by relaxing the usual restrictions on the shape of the $\{\phi_n(\lambda)\}$, considerable advantage can be obtained. There is no theoretical or practical reason, for example, for the $\{\phi_n(\lambda)\}$ to be nonoverlapping. What is needed is to determine the ordered set of basis functions which are optimal with regard to a meaningful system performance criterion.

Illustration

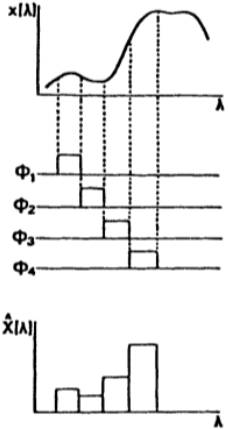


Fig. 7. Approximation of the spectral response function by a set of four basic functions.

Outline

Overall performance

A key consideration is the choice of the optimality criterion. It is desirable to optimize the sensor design with respect to an overall system (including the scene and processor) performance criterion. The probability of correct classification is the criterion to be used here. This choice is made because it is an objective indicator of desired performance in a practical sense for a large proportion of applications, and it is perhaps the best studied and understood in a theoretical sense. In selecting this performance measure there must also be associated with it a processor (classifier) model; in this case we chose the maximum likelihood rule, for the same reasons.

Design factor

However, because of the design factor pointed out in item (2) and Fig. 2, we find it desirable to define an intermediate optimality criterion. Because the sensor must function over a varied collection of strata using any of a large collection of classifiers, a criterion was chosen which is a measure of the fidelity with which the output of the sensor represents the input. We will choose the set $\{\phi_i(\lambda)\}$ such that for a given $x(\lambda)$ the approximation $\hat{x}(\lambda)$ is as close as possible to the true spectral response function. One may think of this approach as one intended to minimize the information loss through the sensor even though it cannot be known to the sensor designer what the information is. In passing from $x(\lambda)$ to $\{x_i\}$ there is no information loss if $x(\lambda)$ is recoverable from $\{x_i\}$.

Criterion

A common criterion for representation accuracy is the expected mean-square representation error given by

$$E\{\epsilon_r\} = E\left\{\int_\Lambda [x(\lambda) - \hat{x}(\lambda)]^2 \, d\lambda\right\}. \tag{3}$$

However, it is desirable at this point to generalize this criterion by introducing a weight function $w(\lambda)$ on the spectral interval. As will be seen, the weight associated with each λ can be used to introduce into the analysis *a priori* knowledge concerning the spectrum. Thus (1) and (3) become [16].

$$x_i = \int_\Lambda x(\lambda)\, \phi_i(\lambda)\, w(\lambda)\, d\lambda \tag{1a}$$

Goal

We want to choose the set of basis functions $\{\phi_i(\lambda)\}$ which is optimal with respect to the spectral representation criterion of expected mean-square error $E\{\epsilon_r\}$. More specifically, it is desired that the representation be complete in the sense that the expected mean-square error for any function in the ensemble be made arbitrarily small simply by including enough terms, that convergence of the approximation to the original response be rapid in the first few terms, and, without loss of generality, we may also ask that the basis functions be orthogonal to each other.

Karhunen-Loeve expansion

A technique for determining the set of optimal basis functions for an ensemble which satisfies the desired properties is based on the weighted Karhunen–Loeve expansion [2], [16], [17]. The solution to the homogeneous linear integral equation is

$$\gamma_i \phi_i(\lambda) = \int_\Lambda K(\lambda, \xi)\, \phi_i(\xi)\, w(\xi)\, d\xi \tag{4}$$

with the covariance function of the stochastic process, $K(\lambda, \xi)$, as kernel is a set of eigenfunctions $\{\phi_i(\lambda)\}$ with corresponding eigenvalues γ_i. If the eigenvalues are arranged in descending order, the corresponding sequence of eigenfunctions can be used to form a linear combination of the eigenfunctions which converges to the original spectral response function with arbi-

Empirical approximation

The optimal sensor design problem may be solved on a digital computer using empirical data taken by field measurements. Some approximations must be made in order to take into consideration some practical constraints. First the response functions are not available as continuous functions but are obtained in the field by sampling the spectrum with an instrument that uses very narrow spectral windows. Secondly, the parameters of the process are not known *a priori*, hence, it is necessary to estimate the mean and covariance functions using a representative sample from the ensemble. Finally, because the data will be stored and processed digitally it is necessary to quantize the amplitude of the response at each of the spectral sample points. Each of these constraints potentially can contribute to the representation error. It has been shown that

Outline

Overall performance

The performance of the overall system is ultimately what we wish to optimize. For this purpose, as previously indicated the probability of correct classification P_c has been chosen as the performance indicator to be optimized. If the vector X is an observation from one of M classes C_i, $i = 1, 2, \cdots, M$ with *a priori* probabilities P_i, the probability of correct classification, using the maximum likelihood rule is given by

$$P_c = \int \max_i \; \{P_i p(X|C_i)\} \, dX \tag{7}$$

where $p(X|C_i)$ is the conditional (multivariant) probability density function for class i. The integral in (7) is over the observation space.

Performance vs representation error

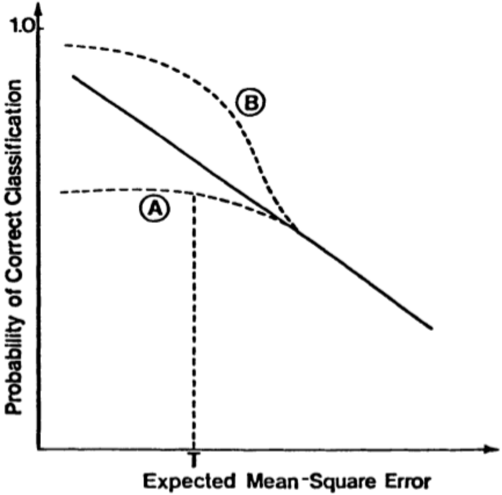


Fig. 8. Probability of a correct classification as a function of expected

Sufficient representation

The addition of terms to the series expansion causes a monotonic decrease in the spectral representation error, but the effect of the additional terms on the overall system performance must be determined. It can be shown that increasing the number of terms in the representation will never decrease the performance provided that the stochastic process is completely known. If after N terms the improvement in performance is small compared to the reduction in representation error, then the representation is sufficient. This is illustrated by case A of Fig. 8 in which the threshold T indicates the minimum required $E[\epsilon_r]$. However, if the performance is showing significant improvement for a small decrease in the mean-square error, case B of Fig. 8, more terms are necessary to complete the representation.

Class related issues

Since the parameters of the stochastic process must be estimated from a sample of the ensemble, the effect of the size of the sample relative to the dimensionality of the system is important. Hughes [10] has shown that if the sample size is too small, the classification performance may actually be degraded by adding terms to the expansion. Thus it is necessary to maintain a large set of sample functions from which to estimate the statistics.

The choice of information classes also influences the performance of the pattern recognition system. For purposes of classifying the data into distinct classes it is required that the class list have the following properties simultaneously [11].

1) Each class must be of interest to the user, i.e., of informational value.

Outline

Outline

Data

A collection of field data consisting of spectral response functions on three dates from Williams County, ND, and three dates from Finney County, KS, was available from the field measurements library at Purdue/LARS. More than one thousand spectra were available from each location and collection date. The response functions were sampled in wavelength using narrow windows of 0.02 μm over the range $0.4 \leqslant \lambda \leqslant 2.4$ μm.

The optimal set of basis functions is found numerically by estimating the covariance matrix from the sample response function. Maximum likelihood estimates of the mean and covariance matrices are given by

$$\bar{X} = E\{X\} \approx \hat{\bar{X}} = \frac{1}{N_s} \sum^{N_s} X_i \qquad (8)$$

Sensors

TABLE I

SPECTRAL BAND LOCATIONS FOR TWO PRACTICAL SENSOR DESIGNS*

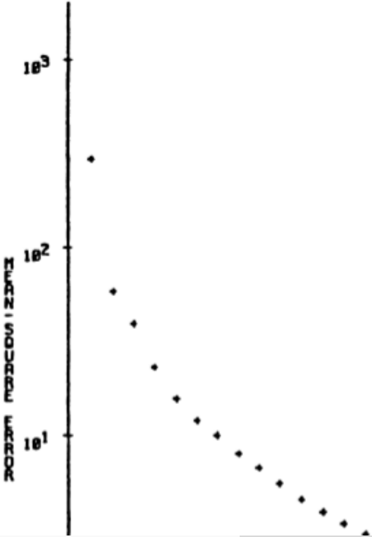
Sensor Number 1					Sensor Number 2				
Band	Wavelength				Band	Wavelength			
k	$\lambda_{\ell k}$	to	λ_{uk}		k	$\lambda_{\ell k}$	to	λ_{uk}	
1	.5	to	.6µm		1	.45	to	.52µm	
2	.6	to	.7µm		2	.52	to	.60µm	
3	.7	to	.8µm		3	.63	to	.69µm	
4	.8	to	1.1µm		4	.76	to	.90µm	
					5	1.55	to	1.75µm	
					6	2.08	to	2.35µm	

*The band edges of these sensors were selected to coincide with the nominal bandwidths of the MSS of Landsats 1, 2, and 3, and those of the TM of Landsat-D.

most of the sensors which were simulated consisted of a small set of rectangular basis

Outline

Representation error vs KL expansion terms



Performance vs representation error