Stochastic model utilizing spectral and spatial characteristics

Hyperspectral session - Landgrebe

Miguel A. Veganzones

Grupo de Inteligencia Computacional
Universidad del País Vasco

2012-03-09

Outline

Outline

Paper

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. PAMI-9, NO. 3, MAY 1987 457

Stochastic Model Utilizing Spectral and Spatial Characteristics

H. M. KALAYEH AND D. A. LANDGREBE

***Abstract*—In remote sensing, because of physical properties of targets, sensor pixels in spatial proximity to one another are class conditionally correlated. Our main objective is to exploit this spatial correlation. Therefore, a two-dimensional causal first order Markov model was used to extract the spatial and spectral information and, based upon it, new object classifiers with improved performance were developed.**

First, the minimum distance (MT) and the maximum likelihood (ML) object classifiers are discussed. Then, based on the proposed model, these two classifiers are modified, and a linear object classifier is introduced. Finally, experimental results are presented.

***Index Terms*—Markov model, maximum likelihood classifier, minimum distance classifier, multispectral image data, object classifier, spatial correlation.**

Motivation

- In remote sensing, because of physical properties of targets, sensor pixels in spatial proximity to one another are class conditionally correlated.
  - Our main objective is to exploit this spatial correlation.
- Therefore, a two-dimensional causal first order Markov model was used to extract the spatial and spectral information and, based upon it, new object classifiers with improved performance were developed.
  - First, the minimum distance (MT) and the maximum likelihood (ML) object classifiers are discussed.
  - Then, based on the proposed model, these two classifiers are modified.
- Data: Landsat MMS (4 bands) and aircraft (12 bands).

Problem definition

Multispectral remotely sensed data consist of an observation set XX, a location set Ω, and a population set C where:

$$XX = X(s),\ s \in \Omega)$$

$$\Omega = \left\{ s = (i, j),\ 1 \leq i \leq I,\ 1 \leq j \leq J \right\}$$

$$\Omega_x = \left\{ s = (i, j),\ I_{1x} \leq i \leq I_{2x},\ J_{1x} \leq j \leq J_{2x} \right\},\ \Omega_x \subseteq \Omega$$

$$C = \left\{ \omega_1,\ \omega_2,\ \cdots,\ \omega_m \right\}$$

and $X(s)$ is a q-dimensional random observation. Let $\{X(s),\ s \in \Omega_x\}$ be the set of observations from an unknown object ω_u. Now, the problem to be addressed is how to classify this object into one of m possible classes.

Spatial correlation

It has been observed in [2]–[7] that two pixels in spatial proximity to one another are class unconditionally and class conditionally correlated. The unconditional correlation usually decays slowly with distance but conditional correlation decreases very rapidly [2]. This spatial correlation can be due to physical properties of the sensor and the target and can also be induced by the atmosphere. Therefore, this spatial correlation introduces redundancy in the object data.

Our objective as mentioned earlier is to 1) extract spatial (class conditional) correlation and 2) generate independent spatial observations for each object. By so doing, additional information useful in class identification/discrimination will be acquired and redundancy will be eliminated.

Outline

Minimum Distance Object (MDO) classifier

Let $p(X(s)|\omega_l)$ denote the class conditional density function for the class ω_l. The decision rule for a minimum distance object classifier [1], [2] is given by:

$$d_{ul} = \min_k d_{uk} \text{ then classify } \{X(s),\, s \in \Omega_x\} \text{ into class } \omega_l$$

where

$$d_{uk} = d\left[P\left(\{X(s),\, s \in \Omega_x\} \middle| \omega_u\right),\, P\left(\{X(s),\, s \in \Omega_y\} \middle| \omega_k\right)\right].$$

Here d_{uk} denotes the statistical distance between the probability density functions of class ω_u and class ω_k. Some popular distance measures such as the Bhattacharyya distance and the divergence for two normal density functions are given in [1], [2]. In our analysis the Bhattacharyya distance was used as a distance measure.

Maximum Likelihood Object (MLO) classifier

The decision rule for the maximum likelihood object classifier [1], [1a] is given by:

$$\text{If } p\left(\left\{X(s),\, s \in \Omega_x\right\} \middle| \omega_l\right) = \max_{k=1,2,\cdots,m} p\left(\left\{X(s),\, s \in \Omega_x\right\} \middle| \omega_k\right) \tag{1}$$

then classify $\{X(s),\, s \in \Omega_x\}$ into class ω_l. Under the assumption [1], [2] that the $X(s)$'s are spatially uncorrelated Gaussian random vectors, the joint class conditional density function can be calculated by

$$p\left(\left\{X)(s),\, s \in \Omega_x\right\} \middle| \omega_l\right) = \prod_{s \in \Omega_x} p\left(X(s) \middle| \omega_l\right). \tag{2}$$

Outline

Outline

Spatial variation model

We assume that spatial variation of observations can be modeled by

$$Y(s) = \sum_{(i,j)\in N} \rho^k_{(i,j)} Y(s + (i, j)) + W(s), \; s \in \Omega_x \qquad (3)$$

where

$$N = \{(0, -1), (-1, -1), (-1, 0)\}.$$

$\rho^k_{(i,j)}$, for $(i, j) \in N$, are spatial correlation of class k and are assumed to be $q \times q$ diagonal matrices and

$$\hat{M}_k = \frac{1}{n_x} \sum_{s\in\Omega_x} X(s)$$

$$Y(s) = X(s) - \hat{M}_k, \; \{X(s), s \in \Omega_x\} \in \omega_k \qquad (4)$$

where $\hat{M}_k$ is the estimated sample mean of class k from location set, Ω_x and N is the neighborhood set.

Noise

Also, $(\{W(s), s \in \Omega_x\} | \omega_k)$, $k = 1, 2, \cdots, m$ are Gaussian white noise field.

$$E\left[W(s) \middle| \omega_k\right] = 0 \tag{5}$$

$$E\left[W(s)\, W^T(t) \middle| \omega_k\right] = \begin{cases} R_k & s = t \\ 0 & s \neq t \end{cases} \tag{6}$$

$$E\left[Y(s) \middle| \left\{Y(s + (i,j)), (i,j) \in N\right\}; \omega_k\right] = \sum_{(i,j) \in N} \rho^k_{(i,j)} Y(s + (i,j))) \tag{7}$$

$$\mathrm{cov}\left[Y(s) \middle| \left\{Y(s + (i,j)), (i,j) \in N\right\}; \omega_k\right] = R_k \tag{8}$$

Observation likelihood

Since $W(s)$'s are spatially uncorrelated Gaussian random vectors,

$$p\left(\left\{W(s),\ s \in \Omega_x\right\} \middle| \omega_k\right) = \prod_{s \in \Omega_x} p\left(W(s) \middle| \omega_k\right). \tag{9}$$

By assumption that the observations are Gaussian and come from a first order Markov process the following can be written:

$$p\left(\left\{Y(s),\ s \in \Omega_x\right\} \middle| \omega_k\right)$$

$$\left[\prod_{s \in \Omega_x} P\left(\left\{Y(s+(i,j)),\ (i,j) \in N\right\} \middle| \omega_k\right)\right]\left[p\left(\left\{Y(s),\ s \in \Omega_b\right\} \middle| \omega_k\right)\right] \tag{10}$$

where

$$\Omega_b = \left\{s(i,j),\ I_{1b} \le i \le I_{2b},\ J_{1b} \le J \le J_{2b}\right\}.$$

Practically, it is not possible to estimate the distributions of the pixels on the boundaries, but if we *do* estimate, because the pixels on the boundaries generally come from a mixture distribution, the second term of (10) will be almost constant. Hence, in computing the decision rule for classifying an object based on the proposed model, this term may be ignored.

Parameter estimation

It is assumed that $\rho^k(i, j)$'s for $(i, j) \neq (0, 0)$ are diagonal matrices. Therefore, the spatial correlation on each channel can be estimated independently of the others. The least square estimate of $\rho^k(i, j)$'s [8]–[10] are given by:

$$\hat{\Theta}_p^k = \left[\prod_{s \in \Omega_x} Z_p(s) \, Z^T(s) \right]^{-1} \left[\sum_{s \in \Omega_x} y_p(s) \, Z_p(s) \right] \qquad (11)$$

where

$$\hat{\Theta}_p^k = \left[\hat{\Theta}_p^k(0, -1), \hat{\Theta}_p^k(-1, 0), \hat{\Theta}_p^k(-1, -1) \right]^T \qquad (12)$$

$$\hat{\Theta}_p^k(i, j) = \left[\rho_1^k(i, j), \rho_2^k(i, j), \cdots, \rho_q^k(i, j) \right]^T$$

$$Z_p(s) = \left[Y_p(s + (0, -1)), Y_p(s + (-1, 0)), \cdot \, Y_p(s + (-1, -1)) \right]^T \qquad (13)$$

$$Y_p(s) = \left[Y_1(s), Y_2(s), \cdots, Y_q(s) \right]'$$

$$p = 1, 2, \cdots, q \text{ and } k = 1, 2, \cdots, m.$$

The estimate of the covariance matrix of $W(s)$ is given by

$$\hat{R}_k = \frac{1}{n_x} \sum_{s \in \Omega_x} \left(Y(s) - \left(\sum_{(i,j) \in N} \hat{\rho}^k(i, j) \, Y(s + (i, j)) \right) \cdot \left(Y(s) - \left(\sum_{(i,j) \in N} \hat{\rho}^k(i, j) \, Y(s) + (i, j) \right) \right)^T \right. . \qquad (14)$$

Outline

Model

Let δ^k be a set of parameters for class ω_k. The existing object classifiers characterize each class or object by two parameters, i.e.,

$$\delta^k = \{\hat{M}_k, \hat{\Sigma}_k\}$$

where $\hat{M}_k$ and $\hat{\Sigma}_k$ are the estimate of the mean vector and covariance matrix of class ω_k. But by the proposed model for each class or object we have

$$\delta^k = \left\{\hat{M}_k, (\hat{\rho}^k(i,j), (i,j) \in N, \hat{R}_k\right\}.$$

The proposed model can be thought of as a filter which maps the data into spatially uncorrelated Gaussian random vectors $W(s)$.

MMDO

The objective here is to modify the MDO classifier so as to increase its effectiveness. The decision rule for the proposed modified minimum distance object classifier is given by (Bhattacharyya distance):

$$B_{ul} = \min_k B_{uk}. \tag{15}$$

Then classify $\{ Y(s),\ s \in \Omega_s \}$ into class ω_l where

$$\hat{B}'_{kl} = \frac{1}{2} \ln \frac{\left| \frac{1}{2}(\hat{R}_k + \hat{R}_k) \right|}{\left| \hat{R}_k \right|^{1/2} \cdot \left| \hat{R}_k \right|^{1/2}}$$

$$\hat{R}_k = \frac{1}{nx} \sum_{s \in \Omega_x} W(s)\, W^T(s)$$

$$W(s) = Y(s) - \sum_{(i,j) \in N} \hat{\rho}^k(i,j)\, Y\big(s + (i,j)\big),\ s \in \Omega_x$$

[see, e.g., (3)], and

$$Y(s) = X(s) - \hat{M}_k.$$

Outline

MLO

If we assume that observations from an object are uncorrelated and Gaussian, then the decision rule for the MLO classifier is given by

$$\text{classify } \{X(s),\ s \in \Omega_x\} \text{ in } \omega_l$$
$$\text{if } \ln p(\{X(s),\ s \in \Omega_x\} \mid \omega_l) = \max_k \ln p(X(s),\ s \in \Omega_x\} \mid \omega_k) \quad (16)$$

where

$$\ln p(\{X(s),\ s \in \Omega_x\} \mid \omega_k)$$
$$= \frac{n_x}{2} [(q \ln 2_n + \ln |\hat{\Sigma}_k|) + \operatorname{tr}(\hat{\Sigma}_k^{-1} \hat{Q}_k))] \quad (17)$$

$$\hat{Q}_k = \frac{1}{n_x} \sum_{s \in \Omega_x} (X(s) - \hat{M}_k)(X(s) - \hat{M}_k)^T. \quad (18)$$

MMLO

But our assumption is that observations in spatial proximity to one another are class conditionally correlated. Based on the proposed model, the decision rule for the modified maximum likelihood object classifier is given by

$$\text{classify } \left\{W(s),\ s \in \Omega\right\} \text{ in } \omega_l$$
$$\text{if } \ln p\left(\left\{W(s),\ s \in \Omega_x\right\} \middle| \omega_l\right) = \max_k \ln p\left(\left\{W(s),\ s \in \Omega_x\right\} \middle| \omega_k\right) \tag{19}$$

where

$$\begin{aligned} \ln p\left(\left\{W(s),\ s \in \Omega_x\right\} \middle| \omega_l\right) &= \ln p\left(\left\{Y(s),\ s \in \Omega_x \middle| \omega_l\right)\right. \\ &= \ln p\left(\left\{X(s),\ s \in \Omega_x\right\} \middle| \omega_l\right) \end{aligned} \tag{20}$$

$$\ln p\left(W(s),\ s \in \Omega_x \middle| \omega_k\right) = -\frac{n_x}{2}\left[\left(q \ln 2\pi + \ln|\hat{R}_k|\right) + \operatorname{tr}\left(\hat{R}_k^{-1}\hat{Q}_k\right)\right] \tag{21}$$

$$\hat{Q}_k = \frac{1}{n_x} \sum_{s \in \Omega_x} W(s)\, W^T(s) \tag{22}$$

$$W(s) = Y(s) - \sum_{(i,j) \in N} \hat{\rho}^k_{(i,j)}\, Y(s + (i,j))$$

and

$$Y(s) = X(s) - \hat{M}_k.$$

Equation (21) clearly shows the dependency of the decision rule on the spatial correlations $\hat{\rho}^k(i,j)$. If $\hat{\rho}^k(i,j)$ are different for inseparable classes, then one expects to see the probability of error in MMLO to be less than the MLO classifier.

Outline

LMDO

When the classes are separable, the data are not complex, or very limited numbers of training samples are available, we do not need (or cannot use) a sophisticated classifier such as MMDO or MMLO. A simple linear minimum distance object classifier (LMDO) can do as well or sometimes even better. This is because in the LMDO classifier, we need only to estimate the mean vector, but in MMDO or MMLO, we must also estimate $\{\rho i, j, (i, j) \in N\}$ and R_k and any inaccuracy in estimates of R_k and $\{\rho i, j, (i, j) \in N\}$ degrades the performance of the MMDO or MMLO [11]. The decision rule for the LMDO classifier is given by

$$L_{uk} = (\hat{M}_u - \hat{M}_k)^T (\hat{M}_u - \hat{M}_k)$$

where

$$\hat{M}_u = \frac{1}{n_x} \frac{\Gamma}{s \in \Omega_x} X(s). \qquad (23)$$

If $L_{ul} = \min_k L_{uk}$, then classify the object into class ω_l.

Outline

Data

- Spatially registered multitemporal Landsat multispectral scanner (MSS) data acquired over Henry County, Indiana, in 1978 and the aircraft data set of the 1971 Corn Blight Watch flightline 210 were selected.

The acquisition dates for the Landsat MMS data are: June 9, July 16, August 20, and September 26, 1978. The classes corn and soybean were chosen for analysis. These two data sets were chosen for analysis for the following reasons:

1) Wall-to-wall ground truth is available. This is important both for deriving good quality training samples and for accurate determination of performance.

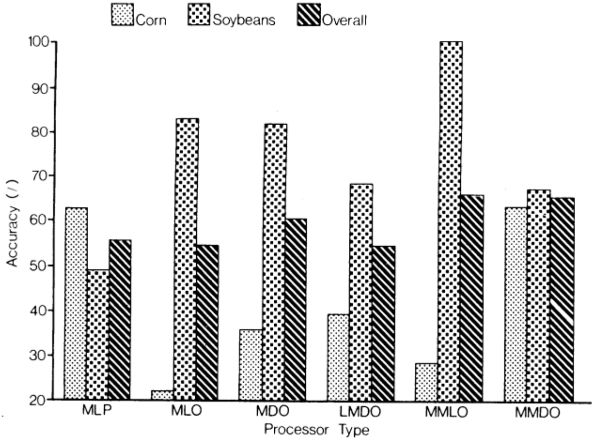
2) The ground spatial resolution of the aircraft data set is much finer than the Landsat MSS data set. It is important to see how this affects the spatial class conditional correlations.

3) The performance of the proposed classifiers with Landsat MSS (4 channels, low ground resolution and 6 bit data representation) and aircraft (12 channels, high ground resolution and 8 bit data representation) data sets under different class separabilities could be evaluated.

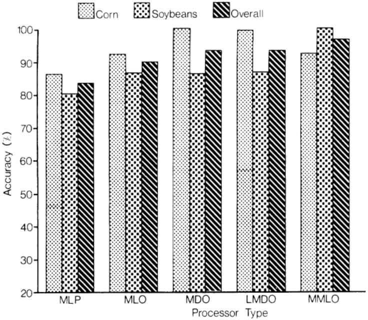
Training methods

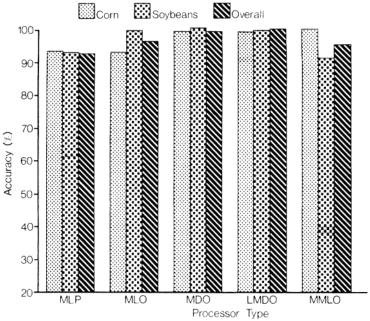
Histogramming and clustering are two commonly used training methods which could be used to find objects with approximately Gaussian observations for training fields. We used the histogramming method to define the spectral classes from training fields or objects. The training objects were chosen to be representative of the informational classes (the size of the smallest object was 10×10 pixels). Then based on a two-dimensional Markov model only horizontal and vertical correlations were extracted. The main reason for this was that horizontal and vertical correlations are more significant than diagonal correlation and furthermore we wanted to keep the complexity of the model as low as possible.

Experiment 1 - Landsat - non separable



Experiment 2,3 - Landsat - separable





Experiment 4 - Aircraft - non separable

TABLE I
CLASSIFICATION PERFORMANCE BY CLASS FOR THREE DIFFERENT CLASSIFIERS (AIRCRAFT DATA)

Maximum Likelihood Pixel Classifier

Group	No. of Samples	Percent Correct	No. of Samples Classified Into WHEAT	No. of Samples Classified Into HAY
1 WHEAT	734	98.5	723	11
2 HAY	862	44.0	483	379
Total	1596	69.0	1206	390

Minimum Distance Object Classifier

Group	No. of Fields	Percent Field Correct	No. of Samples	Percent Samples Correct	No. of Samples Classified Into WHEAT	No. of Samples Classified Into HAY
1 WHEAT	8	100.0	734	100.0	8	0
2 HAY	7	57.1	862	46.4	3	4
Total	15	80.0	1596	71.1	11	4

Modified Minimum Distance Object Classifier

Group	No. of Fields	Percent Field Correct	No. of Samples	Percent Samples Correct	No. of Samples Classified Into WHEAT	No. of Samples Classified Into HAY
1 WHEAT	8	87.5	734	94.6	7	1
2 HAY	7	71.4	862	73.2	2	5
Total	15	80.0	1596	83.0	9	6