

**NEIGHBOURHOOD BASED CLASSIFICATION
OF REMOTELY SENSED DATA
BASED ON GEOMETRIC PROBABILITY MODELS**

by

Nils Lid Hjort

Norwegian Computing Center

and

Stanford University

TECHNICAL REPORT NO. 10

APRIL 1985

PREPARED UNDER THE AUSPICES

OF

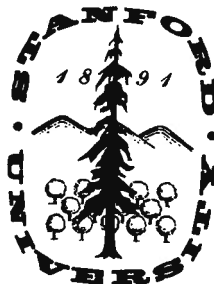
NATIONAL SCIENCE FOUNDATION

GRANT MCS 81-09584

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA



**NEIGHBOURHOOD BASED CLASSIFICATION
OF REMOTELY SENSED DATA
BASED ON GEOMETRIC PROBABILITY MODELS**

**by
Nils Lid Hjort
Norwegian Computing Center
and
Stanford University**

**TECHNICAL REPORT NO. 10
APRIL 1985**

**PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION
GRANT MCS 81-09584**

**DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA**

**NEIGHBOURHOOD BASED CLASSIFICATION
OF REMOTELY SENSED DATA
BASED ON GEOMETRIC PROBABILITY MODELS**

by

Nils Lid Hjort

Norwegian Computing Center

and

Department of Statistics

Stanford University

Abstract

Some methods for classification of picture elements (pixels) of remotely sensed images are considered. The methods are contextual in the sense that measurement vectors from neighbouring pixels are taken into account when the current pixel is classified. The proposed procedures, which use 4 or 8 neighbours, are derived using a simple geometric Poisson field model for the distribution of classes. Parameter estimates can be obtained using training samples or the measurement vectors from the unclassified part of the scene.

1. Introduction.

Consider a scene of pixels. K classes or categories are defined on the scene. Let C_i be the true class of pixel i . Each C_i is in $\{1, \dots, K, \text{OUT}\}$, where OUT includes all pixels not belonging to any of the K predefined classes. Observed in pixel i is a d -dimensional vector X_i , usually having spectral measurements from each of d channels as components. X -vectors from class k follow some density $f_k(x)$.

A statistical classification procedure aims at classifying each pixel, i.e. predicting the C_i 's, on the basis of the observed X -vectors. We assume from now on that the outlier pixels have been taken care of by a separate detection procedure, for example as described in Hjort (1985a). Thus $C_i \in \{1, \dots, K\}$ for the remaining pixels.

We now describe a class of neighbourhood based classification procedures discussed in greater detail in Hjort and Mohn (1984). For convenience and concreteness we consider the case where observations from pixel i and its four immediate neighbours, say

$$\Delta_i = \{X_i, X_{iN}, X_{iE}, X_{iS}, X_{iW}\}, \quad (1)$$

are to be used when classifying pixel i . Let

$$g(a, b, c, d \mid k) = \Pr\{C_{iN} = a, C_{iE} = b, C_{iS} = c, C_{iW} = d \mid C_i = k\} \quad (2)$$

be configuration probabilities for classes, derived from some mosaic process model. Also, let

$$f(\Delta_i \mid k, a, b, c, d) = f_k(X_i)h(\Delta_i \mid k, a, b, c, d) \quad (3)$$

be the simultaneous density of $X_i, X_{iN}, X_{iE}, X_{iS}, X_{iW}$ conditional on the underlying classes being k, a, b, c, d . On the r.h.s. of (3) the individual class density $f_k(X_i)$ is factored out so that $h(\Delta_i \mid k, a, b, c, d)$ is the density of the four neighbour vectors $X_{iN}, X_{iE}, X_{iS}, X_{iW}$ given the class figuration k, a, b, c, d and X_i . (A more appealing but more cumbersome notation would be $h(X_{iN}, X_{iE}, X_{iS}, X_{iW} \mid k, a, b, c, d; X_i)$.) Finally, let $\pi(k)$ denote the so-called a priori probability of class k , interpreted as the proportion of class k pixels among the non-outlying pixels in an imaginary very large scene of which the present scene (or rather the rest of the given scene, after removing pixels from training sets) is a random part; $k = 1, \dots, K$.

Using the quantities introduced above we can consider the important a *posteriori* probability of having class k in a pixel, given the observation vectors from itself and its neighbours:

$$\begin{aligned}
P(k | \Delta_i) &= \Pr\{C_i = k | \Delta_i\} \\
&= \frac{1}{f(\Delta_i)} \pi(k) \sum_{a,b,c,d} g(a,b,c,d | k) f(\Delta_i | k, a, b, c, d) \\
&= \frac{1}{f(\Delta_i)} \pi(k) f_k(X_i) \sum_{a,b,c,d} g(a,b,c,d | k) h(\Delta_i | k, a, b, c, d).
\end{aligned} \tag{4}$$

$f(\Delta_i)$ is the marginal density, making $\sum_{k=1}^K P(k | \Delta_i) = 1$.

Hjort and Mohn (1984) give arguments in favour of using the classification procedure

$$\hat{C}_i = \begin{cases} k & \text{if } P(k | \Delta_i) = \max_{m \leq K} P(m | \Delta_i) > 1 - c, \\ D & \text{if every } P(k | \Delta_i) \leq 1 - c. \end{cases} \tag{5}$$

Here D is a “doubt” option, to be used in the absence of strong evidence, intended for example for pixels being mixtures of points from more than one class. The threshold $1 - c$ originates from a loss function of the type

$$L(C_i, \hat{C}_i) = \begin{cases} 0 & \text{if } \hat{C}_i = C_i \text{ (correct decision),} \\ 1 & \text{if } \hat{C}_i \neq C_i \text{ and } \hat{C}_i \in \{1, \dots, K\} \text{ (wrong decision),} \\ c & \text{if } \hat{C}_i = D \text{ (being in doubt).} \end{cases}$$

If the user wants to avoid the doubt class option, i.e. assign a proper class label to every pixel, he can choose $c \geq 1 - (1/K)$. In this case $\hat{C}_i$ is just the class with highest a posteriori probability.

The classifier (5) is not fully specified in that it requires a statistical model for the behavior of the true classes $\{C_i\}$ on the scene and for the spectral signatures $\{X_i\}$ given the classes, i.e. g and h in (4) must be modelled. Such models will also be defined in terms of unknown parameters, which would then have to be estimated. Hjort and Mohn (1984) discuss several choices for both g and h .

Remark 1. When choosing g and h one should of course strive for good approximations to reality, using prior knowledge about spatial dependence and so on, but attention must also be paid to the resulting complexity of the formula (4): the implementational and computational cost involved in evaluating the large quadruple sum for K classes for each pixel might well

be prohibitive. We should therefore also look for g and h models that simplify or reduce the expression in (4).

The aim of the present paper is to study classifiers of the type (5) for some new choices of g , derived from some simple geometric probability models. As for h in (4) we will limit discussion to the structurally simplest choice that corresponds to assuming X -vectors being conditionally independent given the classes:

$$\begin{aligned} f(\Delta_i | k, a, b, c, d) &= f_k(X_i) f_a(X_{iN}) f_b(X_{iE}) f_c(X_{iS}) f_d(X_{iW}), \\ h(\Delta_i | k, a, b, c, d) &= f_a(X_{iN}) f_b(X_{iE}) f_c(X_{iS}) f_d(X_{iW}). \end{aligned} \tag{6}$$

This has been a common assumption in the literature, see for example Welch and Salter (1971), Swain, Vardeman, and Tilton (1981), Haslett (1983), Owen (1984). The class densities themselves are typically taken to be Gaussian, with identical, proportional, or different covariance matrices.

Remark 2. Hjort, Mohn, and Storvik (1985) argue that the conditional independence hypothesis becomes less adequate for high resolution satellites, i.e. when the pixel size decreases. They propose a h -model based on an autocorrelated noise process. The classifiers presented in this paper may be generalised by using the h -model with the mentioned spatial dependence property, see in particular “specification 3” of Hjort, Mohn, and Storvik (1985), at the price of added complexity.

The paper of Owen (1984) is similar in spirit to the present work. He assumes that Nature employs a particular Poisson field mechanism devised by Switzer (1965) when dividing the landscape into categories, and obtains formulae for the $g(a, b, c, d | k)$'s in terms of an intensity parameter β . His method of estimating β requires known “ground truth” along (at least) a row of pixels.

Owen's work is here extended in several directions. Section 2 relaxes the above mentioned geometric probability model. Owen (1984, p. 195) notes that when data are noisy and the underlying categories shift smoothly accuracy could be increased if information from more than the “pixel cross” with 5 pixels were used. In Section 3 the case of “pixel boxes” with 9 pixels is considered. Finally procedures for obtaining the necessary parameter estimates are given in Section 4. Included is a method that uses all the unclassified vectors on the scene to

obtain estimates for the "contextual parameters" in g , assuming training sets are available to get estimates of class densities first.

The classification procedures described in the present paper are currently being implemented, on a VAX 11/780 tied to a RAMTEK colour screen, at the Norwegian Computing Center. The methods are being tried out on simulated SPOT-scenes and on artificially generated data. Some results are reported in Hjort and Mohn (1984) and in the project report by Sæbø et al. (1985). The experience to date, though limited, strongly indicates that the contextual methods can offer much improvement over the traditional non-contextual ones that are currently used in commercial systems.

2. A Simple Geometric Probability Model.

2.1. The p, q, r Model

Assume that Nature divides the scene under study into a set of regions $R, S, \dots$. Assume further that the average (random) region is large compared with the pixel size, so that most pixel crosses will have points from at most two regions. This motivates admitting positive probability to only the three types of crosses below:

R			R			R		
R	R	R	S	R	R	R	R	R
R			S			S		

These patterns are denoted X , L , and T respectively. There are 4 possible L patterns and 4 possible T patterns. Suppose that the probability is p of getting an X , $q/4$ of getting one of the 4 L 's, and $r/4$ of getting one of the 4 T 's; $p + q + r = 1$.

Next, Nature assigns class labels to these regions, class label k with probability $\pi(k)$, independently for each region. (In particular, adjacent regions may be assigned to the same class.) A pixel is given the class label of its centre point. It follows that $\Pr\{C_i = k\} = \pi(k)$,

and using the notation in (2) we get

$$\begin{aligned}\pi(k)g(k, k, k, k | k) &= p\pi(k) + \frac{1}{4}q\pi(k)^2 + \dots + \frac{1}{4}r\pi(k)^2 \\ &= \pi(k)\{p + (q + r)\pi(k)\}, \\ \pi(k)g(k, k, m, m | k) &= \frac{1}{4}q\pi(k)\pi(m), \\ \pi(k)g(k, k, m, k | k) &= \frac{1}{4}r\pi(k)\pi(m),\end{aligned}$$

$m \neq k$. Hence

$$\left. \begin{aligned}g(k, k, k, k | k) &= p + (q + r)\pi(k), \\ g(k, k, m, m | k) &= g(m, k, k, m | k) = g(m, m, k, k | k) \\ &= g(k, m, m, k | k) = \frac{1}{4}q\pi(m), \\ g(k, k, m, k | k) &= g(k, k, k, m | k) = g(m, k, k, k | k) \\ &= g(k, m, k, k | k) = \frac{1}{4}r\pi(m),\end{aligned} \right\} \quad (7)$$

$m \neq k$. Only these $8K - 7$ terms are positive, making a substantial reduction of the quadruple sum in the general formula (4).

The g -specification above determines, together with the h -specification (6), a contextual classification procedure. According to (4) and (5) we are to find the class k having highest value of

$$\pi(k)f_k(X_i)R_k(\Delta_i),$$

where $R_k(\Delta_i)$ acts as a "spatial correction factor" to the standard noncontextual procedure, and is given by

$$\begin{aligned}R_k(\Delta_i) &= \sum_{a,b,c,d} g(a, b, c, d | k) f_a(X_{iN}) f_b(X_{iE}) f_c(X_{iS}) f_d(X_{iW}) \\ &= \{p + (q + r)\pi(k)\} f_k(X_{iN}) f_k(X_{iE}) f_k(X_{iS}) f_k(X_{iW}) \\ &\quad + \frac{1}{4}q \sum_{m \neq k} \pi(m) f_k(X_{iN}) f_k(X_{iE}) f_m(X_{iS}) f_m(X_{iW}) \\ &\quad + \dots \\ &\quad + \frac{1}{4}r \sum_{m \neq k} \pi(m) f_k(X_{iN}) f_m(X_{iE}) f_k(X_{iS}) f_k(X_{iW}).\end{aligned}$$

This can be written

$$R_k(\Delta_i) = pA_k(\Delta_i) + qB_k(\Delta_i) + rC_k(\Delta_i) \quad (8)$$

after some manipulation, where

$$\begin{aligned}
A_k(\Delta_i) &= f_k(X_{iN})f_k(X_{iE})f_k(X_{iS})f_k(X_{iW}), \\
B_k(\Delta_i) &= \frac{1}{4}[f_k(X_{iN})f_k(X_{iE})b(X_{iS}, X_{iW}) + f_k(X_{iE})f_k(X_{iS})b(X_{iW}, X_{iN}) \\
&\quad + f_k(X_{iS})f_k(X_{iW})b(X_{iN}, X_{iE}) + f_k(X_{iW})f_k(X_{iN})b(X_{iE}, X_{iS})], \\
C_k(\Delta_i) &= \frac{1}{4}[f_k(X_{iN})f_k(X_{iE})f_k(X_{iS})a(X_{iW}) + f_k(X_{iN})f_k(X_{iE})a(X_{iS})f_k(X_{iW}) \\
&\quad + f_k(X_{iN})a(X_{iE})f_k(X_{iS})f_k(X_{iW}) + a(X_{iN})f_k(X_{iE})f_k(X_{iS})f_k(X_{iW})],
\end{aligned}$$

and where

$$\begin{aligned}
a(x) &= \sum_{m=1}^K \pi(m)f_m(x), \\
b(x, y) &= \sum_{m=1}^K \pi(m)f_m(x)f_m(y).
\end{aligned} \tag{9}$$

The resulting classifier may thus be described as follows: find the class k giving highest value of $\pi(k)f_k(X_i)R_k(\Delta_i)$, with $R_k(\Delta_i)$ calculated as above, and let $\hat{C}_i$ be this k , unless

$$\pi(k)f_k(X_i)R_k(\Delta_i) \leq (1 - c) \sum_{m=1}^K \pi(m)f_m(X_i)R_m(\Delta_i),$$

in which case the doubt reject option should be used.

Of course the parameters p , q , r , and $\pi(k)$ and $f_k(x)$, need to be estimated. Procedures managing this are in Section 4.

2.2. The Poisson Field Model

Owen (1984) assumes that the regions $R, S, \dots$ come from a partitioning of the plane by straight lines that are generated according to a Poisson field model introduced by Switzer (1965). If the intensity of this Poisson field is small (relative to the pixel size) then the probability of having more than one line intersecting a pixel cross is negligible. Let β be the probability that exactly one boundary line intersects a given pixel cross neighbourhood. Owen shows that

$$p = 1 - \beta, \quad q = (\sqrt{2} - 1)\beta, \quad r = (2 - \sqrt{2})\beta \tag{10}$$

(in our notation).

The p, q, r model is only slightly more general than Owen's (1984), having two unknown parameters instead of one. Not much is lost in efficiency if the more flexible p, q, r model is used and (10) happens to be true.

Since our quest is the construction of good classifiers we are not very much concerned with probabilistic properties of the underlying stochastic processes, in particular we don't engage ourselves in the question of whether there really exists a stationary mosaic process on a lattice with configuration probabilities (7). The p, q, r model is less dependent on modelling assumptions, and in view of Remark 1.1 we consider (7), and the later fitting of data to these, only as reasonable simplifying equations.

We remark also that it is easy to generalise the classifier described after (9) in several ways. One possibility is to let the p, q, r parameters be class-dependent (i.e. use p_k, q_k, r_k in (8)); another is to let them be able to change smoothly over the landscape.

3. Classifiers Based on 8 Neighbours.

The present section will extend both the p, q, r model and its special case, the Poisson field model that yielded (10), to the case where 8 neighbours of the current pixel are taken into account, i.e. data from the full 3×3 "pixel box" is used to classify its centre pixel.

3.1. The p, q, r, s, t, u Model

It is easy to give a formula for the a posteriori probability of having class k in pixel i , given observation vectors from the pixel and its 8 nearest neighbours, analogous to (4)

$$\begin{aligned} P(k | \Delta_i) &= \Pr\{C_i = k | X_i, X_{i,1}, \dots, X_{i,8}\} \\ &= \frac{1}{f(\Delta_i)} \pi(k) \sum_{a_1, \dots, a_8} g(a_1, \dots, a_8 | k) f(\Delta_i | k, a_1, \dots, a_8) \\ &= \frac{1}{f(\Delta_i)} \pi(k) f_k(X_i) \sum_{a_1, \dots, a_8} g(a_1, \dots, a_8 | k) h(\Delta_i | k, a_1, \dots, a_8). \end{aligned} \quad (11)$$

Here

$$g(a_1, \dots, a_8 | k) = \Pr\{C_{i,1} = a_1, \dots, C_{i,8} = a_8 | C_i = k\} \quad (12)$$

are configuration probabilities, and $(C_{i,1}, X_{i,1}), \dots, (C_{i,8}, X_{i,8})$ denote (true class, observation vector) for pixels counted clockwise from the upper left hand (northwestern) corner of the pixel

box having centre pixel i . Furthermore,

$$\Delta_i = \{X_i, X_{i,1}, \dots, X_{i,s}\}, \quad (13)$$

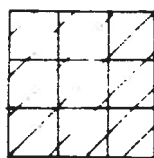
and f and h in (11) have the obvious meanings, parallelling the situation in the previous section. For simplicity we don't change the notation of Δ_i ; it should be clear from the context whether 4 or 8 neighbours are used. This remark also applies to some of the notation used below.

Recipe (5), along with specifications of g and h in (11), provides a contextual classification procedure. However, the sum in (11) contains the potentially enormous number K^6 of terms, and must be reduced or simplified in order to meet practical restrictions.

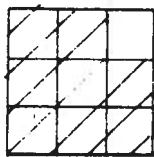
We may reason as in Section 2. Assume again that randomly constructed regions are assigned class labels $1, \dots, K$ with probabilities $\pi(1), \dots, \pi(K)$, independently of each other. A pixel is given the class label of its centre point. Suppose that the regions are large compared with pixel size so that a given pixel box will contain points from more than two regions with only negligible probability. Thus it is not unreasonable to assume that at most one boundary curve will intersect a given pixel box and that such an intersection segment will be approximately linear. It follows from these assumptions that only six different types of patterns of regions should have positive probability. These are denoted A, B, C, D, E , and F , and are displayed in Figure 1. (Displayed in Figure 1 are really the *pixel* patterns that result from underlying region patterns by letting a pixel's region be that of its centre point.)

To construct a (possibly crude, but possibly effective) statistical model for the configuration probabilities (12), postulate that as a result of Nature's random division of the landscape into regions, a given neighbourhood has probability p of being given an A pattern, and similarly probability $q/4, r/8, s/4, t/4, u/8$ of being given respectively a particular B , a particular C , a particular D , a particular E , and a particular F pattern; $p + q + r + s + t + u = 1$. Then

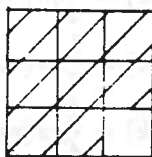
Figure 1



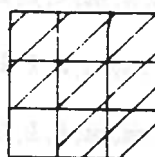
A



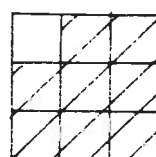
B₁



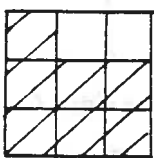
B₂



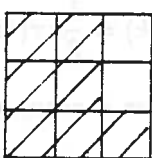
B₃



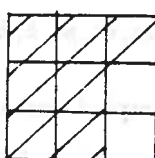
B₄



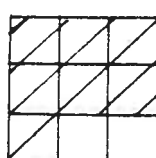
C₁



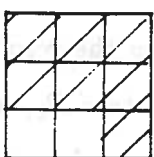
C₂



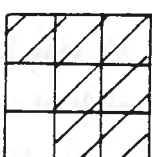
C₃



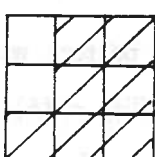
C₄



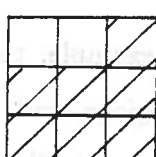
C₅



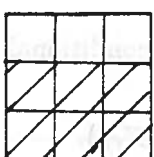
C₆



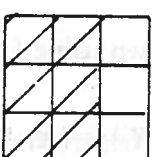
C₇



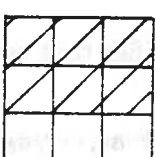
C₈



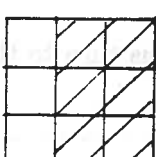
D₁



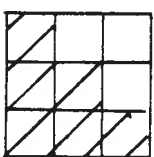
D₂



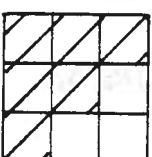
D₃



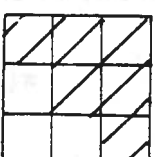
D₄



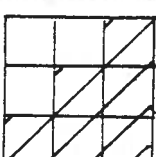
E₁



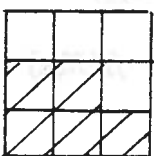
E₂



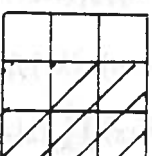
E₃



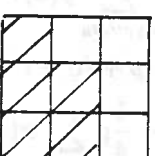
E₄



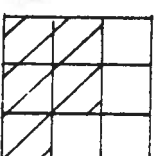
F₁



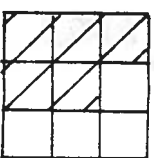
F₂



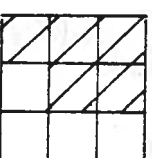
F₃



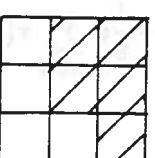
F₄



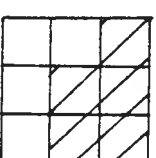
F₅



F₆



F₇



F₈

we may derive, analogously to (7) in the pixel cross case,

$$\left. \begin{aligned} g(k, k, k, k, k, k, k, k | k) &= p + (1 - p)\pi(k), \\ g(k, k, m, k, k, k, k, k | k) &= \frac{1}{4}q\pi(m), \\ g(k, m, m, k, k, k, k, k | k) &= \frac{1}{8}r\pi(m), \\ g(m, m, m, k, k, k, k, k | k) &= \frac{1}{4}s\pi(m), \\ g(k, m, m, m, k, k, k, k | k) &= \frac{1}{4}t\pi(m), \\ g(m, m, m, m, k, k, k, k | k) &= \frac{1}{8}u\pi(m), \end{aligned} \right\} \quad (14)$$

$m \neq k$. Here we have only sampled one of the four B cases, one of the eight C cases, and so on; a full listing has 29 entries, corresponding to the patterns shown in Figure 1. Only $28K - 27$ of the K^8 probabilities in (12) are positive.

For example, the pixel pattern with nine k 's may have been the result of an A pattern of regions, with probability $p\pi(k)$, or the result of region pattern B_1 , with probability $\frac{1}{4}q\pi(k)^2$, etc. Adding these together and dividing out $\pi(k)$ we obtain the conditional probability $g(k, k, k, k, k, k, k, k | k)$ given above.

Let us turn to the classifier that results when using (14) and conditional independence,

$$f(\Delta_i | k, a_1, \dots, a_8) = f_k(X_i) f_{a_1}(X_{i,1}) \cdots f_{a_8}(X_{i,8}).$$

The a posteriori probabilities (11) are proportional to

$$\pi(k) f_k(X_i) R_k(\Delta_i),$$

where the "contextual correction factor" is

$$\begin{aligned} R_k(\Delta_i) &= \sum_{a_1, \dots, a_8} g(a_1, \dots, a_8 | k) f_{a_1}(X_{i,1}) \cdots f_{a_8}(X_{i,8}) \\ &= \{p + (q + r + s + t + u)\pi(k)\} f_k(X_{i,1}) \cdots f_k(X_{i,8}) \\ &\quad + \frac{1}{4}q \sum_{m \neq k} \pi(m) f_m(X_{i,3}) \prod_{j \neq 3} f_k(X_{i,j}) \\ &\quad + \cdots \\ &\quad + \frac{1}{8}u \sum_{m \neq k} \pi(m) \prod_{j=1,2,7,8} f_m(X_{i,j}) \prod_{j=3,4,5,6} f_k(X_{i,j}), \end{aligned}$$

corresponding to the 29 entries in the table that (14) samples. This can be written in the form

$$R_k(\Delta_i) = \prod_{j=1}^8 f_k(X_{i,j}) \{pA_k(\Delta_i) + qB_k(\Delta_i) + rC_k(\Delta_i) + sD_k(\Delta_i) + tE_k(\Delta_i) + uF_k(\Delta_i)\}, \quad (15)$$

where

$$\begin{aligned} A_k(\Delta_i) &= 1 \\ B_k(\Delta_i) &= \frac{1}{4} \sum_{1,3,5,7} a_k(X_{i,j}), \\ C_k(\Delta_i) &= \frac{1}{8} \sum_{(1,2),(2,3),(3,4),(4,5),(5,6),(6,7),(7,8),(8,1)} b_k(X_{i,j}, X_{i,l}), \\ D_k(\Delta_i) &= \frac{1}{4} \sum_{(1,2,3),(3,4,5),(5,6,7),(7,8,1)} c_k(X_{i,j}, X_{i,l}, X_{i,m}), \\ E_k(\Delta_i) &= \frac{1}{4} \sum_{(2,3,4),(4,5,6),(6,7,8),(8,1,2)} c_k(X_{i,j}, X_{i,l}, X_{i,m}), \\ F_k(\Delta_i) &= \frac{1}{8} \sum_{\substack{(1,2,3,4),(2,3,4,5),(3,4,5,6),(4,5,6,7) \\ (5,6,7,8),(6,7,8,1),(7,8,1,2),(8,1,2,3)}} d_k(X_{i,j}, X_{i,l}, X_{i,m}, X_{i,n}), \end{aligned}$$

and where

$$\begin{aligned} a_k(x) &= \sum_{m=1}^K \pi(m) \frac{f_m(x)}{f_k(x)}, \\ b_k(x, y) &= \sum_{m=1}^K \pi(m) \frac{f_m(x) f_m(y)}{f_k(x) f_k(y)}, \\ c_k(x, y, z) &= \sum_{m=1}^K \pi(m) \frac{f_m(x) f_m(y) f_m(z)}{f_k(x) f_k(y) f_k(z)}, \\ d_k(x, y, z, v) &= \sum_{m=1}^K \pi(m) \frac{f_m(x) f_m(y) f_m(z) f_m(v)}{f_k(x) f_k(y) f_k(z) f_k(v)}. \end{aligned}$$

The classifier picks for a given pixel i the class k for which $\pi(k) f_k(\Delta_i) R_k(\Delta_i)$ is highest, unless

$$\pi(k) f_k(\Delta_i) R_k(\Delta_i) \leq (1 - c) \sum_{m=1}^K \pi(m) f_m(\Delta_i) R_m(\Delta_i),$$

in which case one resorts to the doubt class.

3.2. The Poisson Field Model with 8 Neighbours

We now derive formulae for p, q, r, s, t, u of (14) in terms of a single parameter α , assuming Nature to employ Switzer's Poisson field model in her division of the plane into regions,

mentioned in Section 2.2. Thus we obtain the extension to the 8-neighbour case that Owen (1984, p. 195) seems to ask for. The reasoning is similar to Owen's Appendix I, but necessarily becomes more involved and cumbersome.

The Poisson field can be described as follows. Generate points P_i with polar coordinates (r_i, θ_i) in a large disc around an origin O containing the scene of interest. The r_i 's are from a Poisson process in $(0, \infty)$ and the θ_i 's are independent of these and i.i.d. uniforms on $(0, 2\pi)$. Construct for each P_i the line ℓ_i through P_i perpendicular to OP_i . These lines are the boundaries of the induced regions.

Solomon (1978) shows that the number of lines of the field that intersect a given line segment of length d has a Poisson distribution with parameter τd , $\tau > 0$ being a constant characterising the field. Let N_B, N_C, N_D, N_E, N_F be the number of lines that intersect a given pixel box and thereby induce respectively B, C, D, E , and F patterns. (We have $N_B + \dots + N_F \leq 1$ with high probability according to assumptions.) These are independent and Poisson distributed with means $\lambda_B, \lambda_C, \lambda_D, \lambda_E, \lambda_F$, say, by the arguments given in Owen (1984, p. 197).

Consider the pixel box neighbourhood in Figure 2. Take the pixel size to be 1×1 . Patient examination shows that a Poisson field generated line intersecting the segment of unit length shown in Figure 2(i) can induce one of the patterns $C_1, C_3, D_1, E_1, E_4, F_1, F_2, F_3, F_8$, and none of the others displayed in Figure 1. Hence

$$\tau = 2c + d + 2e + 4f$$

where

$$b = \frac{1}{4}\lambda_B, \quad c = \frac{1}{8}\lambda_C, \quad d = \frac{1}{4}\lambda_D, \quad e = \frac{1}{4}\lambda_E, \quad f = \frac{1}{8}\lambda_F, \quad (16)$$

using a rotational invariance argument (the variables $N_{F_1}, \dots, N_{F_8}$ that add up to N_F must have the same distribution, for example).

Summoning the needed perseverance we may similarly carefully enumerate which of the 29 patterns of Figure 1 that can be induced by having a straight line intersect the line segments shown in Figure 2 (ii), (iii), (iv), (v). Intersecting the line segment of length $\sqrt{2}$ in (ii), for

Figure 2

Listed are in each case the different patterns that can be induced when a straight line intersects the indicated line segment.

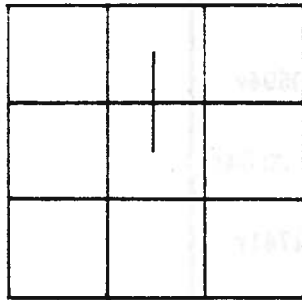
Fig. 2(i): $C_1 \ C_8 \ D_1 \ E_1 \ E_4 \ F_1 \ F_2 \ F_3 \ F_8$

Fig. 2(ii): $B_4 \ C_7 \ C_8 \ D_1 \ D_4 \ E_4 \ F_1 \ F_2 \ F_7 \ F_8$

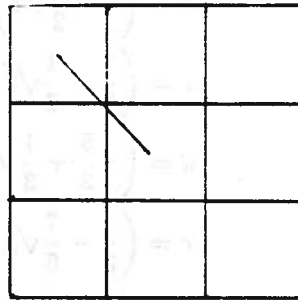
Fig. 2(iii): $B_4 \ C_1 \ C_7 \ D_4 \ E_1 \ F_3 \ F_7$

Fig. 2(iv): $C_1 \ C_6 \ C_7 \ C_8 \ D_1 \ D_4 \ E_1 \ E_3 \ F_1 \ F_3 \ F_6 \ F_7$

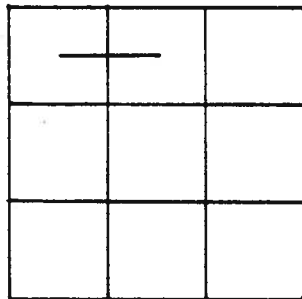
Fig. 2(v): $B_4 \ C_2 \ C_3 \ C_8 \ D_1 \ D_2 \ D_4 \ E_1 \ E_2 \ E_4 \ F_2 \ F_3 \ F_4 \ F_5 \ F_7 \ F_8$



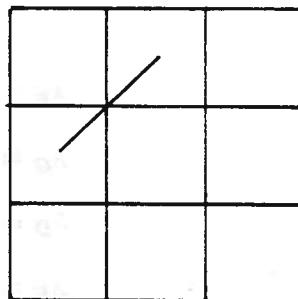
(i)



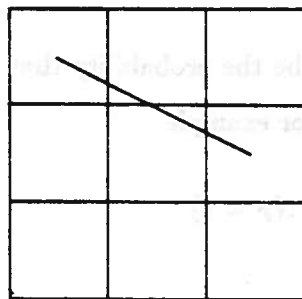
(ii)



(iii)



(iv)



(v)

example, may induce the 10 different patterns listed there, and so on. We end up with equations

$$\left. \begin{aligned} 2c + d + 2e + 4f &= r \\ b + 2c + 2d + e + 4f &= \sqrt{2}r \\ b + 2c + d + e + 2f &= r \\ 4c + 2d + 2e + 4f &= \sqrt{2}r \\ b + 3c + 3d + 3e + 6f &= \sqrt{5}r \end{aligned} \right\} \quad (17)$$

These can be solved to give

$$\left. \begin{aligned} b &= \left(1 - \frac{1}{2}\sqrt{2}\right) r = .2929r \\ c &= \left(\frac{1}{3} + \frac{1}{3}\sqrt{2} - \frac{1}{3}\sqrt{5}\right) r = .0594r \\ d &= \left(-\frac{5}{3} + \frac{1}{3}\sqrt{2} + \frac{2}{3}\sqrt{5}\right) r = .2954r \\ e &= \left(\frac{1}{3} - \frac{7}{6}\sqrt{2} + \frac{2}{3}\sqrt{5}\right) r = .1741r \\ f &= \left(\frac{1}{3} + \frac{1}{3}\sqrt{2} - \frac{1}{3}\sqrt{5}\right) r = .0594r \end{aligned} \right\} \quad (18)$$

and hence

$$\lambda_B = 4b = 1.1716r$$

$$\lambda_C = 8c = .4751r$$

$$\lambda_D = 4d = 1.1818r$$

$$\lambda_E = 4e = .6965r$$

$$\lambda_F = 8f = .4751r$$

Now we can obtain formulae for p, q, r, s, t, u . Let α be the probability that a given pixel box is intersected by a Poisson field line. Then we have for example

$$\begin{aligned} q &= \alpha \Pr\{N_B = 1 \mid N_B + \dots + N_F = 1\} \\ &= \alpha \lambda_B / (\lambda_B + \dots + \lambda_F) \\ &= \frac{1}{4} \alpha \lambda_B / r = \alpha b / r, \end{aligned}$$

and similarly for r, s, t, u . The result is

$$\left. \begin{aligned} p &= 1 - \alpha \\ q &= \alpha b / \tau = .2929\alpha \\ r &= 2\alpha c / \tau = .1188\alpha \\ s &= \alpha d / \tau = .2954\alpha \\ t &= \alpha e / \tau = .1741\alpha \\ u &= 2\alpha f / \tau = .1188\alpha \end{aligned} \right\} \quad (19)$$

The classifier described at the end of Section 3.1 works under the Poisson field modelling assumption by inserting the above found values for p, q, r, s, t, u in (15). The computational complexity is the same. (15) can now be written

$$\begin{aligned} R_k(\Delta_i) &= \prod_{j=1}^8 f_k(X_{i,j}) \{1 + \alpha[.2929B_k(\Delta_i) \\ &\quad + .1188C_k(\Delta_i) + .2954D_k(\Delta_i) + .1741E_k(\Delta_i) + .1188F_k(\Delta_i) - 1]\}. \end{aligned} \quad (20)$$

The problem of obtaining an estimate of α is studied in the next section.

It is interesting to note that formula (10), which was Owen's (1984) result for the 4-neighbour case, follows from the results of the present section. To see this, notice from Figure 1 that A and B patterns produce the X pattern for the pixel cross, and similarly that C and D give T patterns, and E and F give L patterns. Hence

$$\Pr(X) = p + q, \quad \Pr(T) = r + s, \quad \Pr(L) = t + u.$$

The results obtained for p, q, r, s, t, u give

$$\begin{aligned} \Pr(X) &= 1 - \alpha + \alpha b / \tau = 1 - \frac{1}{\sqrt{2}}\alpha, \\ \Pr(T) &= \alpha \left(\frac{2c}{\tau} + \frac{d}{\tau} \right) = \alpha(\sqrt{2} - 1) = \frac{\alpha}{\sqrt{2}}(2 - \sqrt{2}), \\ \Pr(L) &= \alpha \left(\frac{e}{\tau} + \frac{2f}{\tau} \right) = \alpha \left(1 - \frac{1}{2}\sqrt{2} \right) = \frac{\alpha}{\sqrt{2}}(\sqrt{2} - 1). \end{aligned}$$

These are exactly the equations of (10) with $\beta = \alpha/\sqrt{2}$.

4. Parameter Estimates.

In order to put the classifiers described in previous sections to work one needs input values for the unknown parameters involved. Needed in any case are estimates of the a priori probabilities $\pi(k)$ and the class densities $f_k(x)$. The pixel cross method uses either p, q, r or the single parameter β , whereas the pixel box method uses either p, q, r, s, t, u or the single parameter α . The present section gives some methods of estimating these parameters.

4.1. Class Densities and A Priori Probabilities

The tradition of taking the individual class densities to be Gaussian has been the dominant one in remote sensing literature (this assumption is often made implicitly). Exceptions are cases where a class, maybe after further inspection, is seen to have distinct "hidden subclasses", corresponding to more than one bump in histograms. In such cases the $f_k(x)$ in question may be modelled as a mixture of Gaussian components.

Usually a training set is available for each class, i.e. there are observation vectors $X_1^{(k)}, \dots, X_{n_k}^{(k)}$ from pixels that are known to belong to class k ; $k = 1, \dots, K$. These training sets are used in the usual way to obtain estimates for parameters in the class densities $f_k(x)$.

Depending on sample sizes and other factors one may choose among (at least) three possible models:

(i) $f_k(x) = N_d(\mu_k, \Sigma)(x)$,

(ii) $f_k(x) = N_d(\mu_k, \lambda_k \Sigma)(x)$,

(iii) $f_k(x) = N_d(\mu_k, \Sigma_k)(x)$.

In case (iii) one uses the standard empirical mean and covariance matrix for each class:

$$\hat{\mu}_k = \frac{1}{n_k} \sum_{j=1}^{n_k} X_j^{(k)}, \quad \hat{\Sigma}_k = \frac{1}{n_k} \sum_{j=1}^{n_k} (X_j^{(k)} - \hat{\mu}_k)(X_j^{(k)} - \hat{\mu}_k)'$$

In case (i) one keeps $\hat{\mu}_k$ and pools the matrices to get $\hat{\Sigma} = \sum_{k=1}^K (n_k/n) \hat{\Sigma}_k$; $n = \sum_{k=1}^K n_k$. In the interesting intermediate case (ii) one may use one of several possible iteration procedures to obtain the maximum likelihood solutions $\tilde{\lambda}_k, \tilde{\Sigma}$. Owen (1984) uses this model of proportional covariance matrices and gives such a computational procedure. He notes (op. cit., Appendix

II) that it is not known whether the likelihood equations he solves has a unique solution. It has recently been proved, however, that there always is one and only one solution (Søren Johansen (1985), personal communication).

The a priori probabilities $\pi(k)$ are sometimes taken to be equal, i.e. $\pi(k) \equiv \frac{1}{K}$, or perhaps crudely assessed from past experience. It is perfectly possible to estimate the $\pi(k)$'s on the given scene, however, using vectors from all unclassified pixels, provided training sets are available as above, so that initial class density estimates can be obtained. This is pointed out in Hjort and Mohn (1984) and elaborated in Hjort (1985b). The latter reference also give methods by which it is possible to "update" class description parameter estimates, say $\hat{\mu}_k$, $\hat{\Sigma}_k$, for the given scene, again using the unclassified vectors.

4.2. Estimating Contextual Parameters in the 4-Neighbours Case

We must construct estimates of p, q, r and of β in the model given by (7) and (10). There are in fact several methods, depending also on the form in which relevant data is available.

Assume first that "ground truth" has been obtained for M sampled pixel crosses. These could, perhaps ideally, constitute a random sample, or they could be available as a result of examining three (or more) adjacent parallel rows of pixels in the terrain. The M crosses should not be taken from the training sets of the classes, however, if these are chosen as they usually are, as homogeneous regions lying safely within the appropriate classes.

Let M_X be the number of crosses among the M that have all 5 pixels from the same class. Similarly, let M_L and M_T count the L and T patterns observed among the M , as defined in Section 2.1. Let us discard all other crosses, and redefine M if necessary, so that $M_X + M_L + M_T = M$. Then M_X/M is an estimate of

$$\sum_{k=1}^K \pi(k) g(k, k, k, k | k) = \sum_{k=1}^K \pi(k) \{p + (1-p)\pi(k)\} \\ = p + (1-p)w,$$

where

$$w = \sum_{k=1}^K \pi(k)^2. \quad (21)$$

Similarly, M_L/M is close to $\sum_k \sum_{m \neq k} \pi(k) q \pi(m) = q(1 - w)$ and M_T/M is close to $r(1 - w)$.

Thus natural estimators are

$$\begin{aligned} \hat{q} &= \frac{M_L/M}{1 - \hat{w}}, \quad \hat{r} = \frac{M_T/M}{1 - \hat{w}}, \\ \hat{p} &= 1 - \hat{q} - \hat{r} = \frac{M_X/M - \hat{w}}{1 - \hat{w}}, \end{aligned} \quad (22)$$

where $\hat{w} = \sum_{k=1}^K \hat{\pi}(k)^2$ estimates w .

The form of these estimators are not dependent upon whether the M pixels were sampled randomly or systematically, but their properties (standard deviations, for example) do depend on the sampling procedure. It is easy to show that if the M crosses were chosen randomly from a large scene (with (7) meaning stationary probabilities), so that (M_X, M_L, M_T) is multinomial, then $\hat{p}, \hat{q}, \hat{r}$ above are identical to the maximum likelihood solutions.

Estimators of β can also be given based on $M_X/M, M_L/M, M_T/M$. One possibility is to solve the equation

$$M_X/M \doteq p + (1 - p)w = 1 - \beta + \beta w$$

for β , i.e.

$$\hat{\beta} = \left(1 - \frac{M_X}{M}\right) / (1 - \hat{w}) = \hat{q} + \hat{r}, \quad (23)$$

cf. (10). This also happens to be the maximum likelihood solution when (M_X, M_L, M_T) is assumed multinomial.

Owen (1984) gives a similar method when ground truth is available only in nonadjacent parallel rows of pixels.

The methods above require the user to have ground truth for a representative set of pixels outside the individual class training sets. Obtaining such information is most often a demanding and sometimes impossible task, however.

It is fortunate, then, that it is still possible to estimate p, q, r , and β , as we now show. One can in fact use all unclassified vectors to perform this task, without having to examine the terrain outside the class training sets. The reason why this is possible is that data Δ_i from a pixel cross of unknown land, as in (1), has density

$$f(\Delta_i) = \sum_{k=1}^K \pi(k) f_k(X_i) R_k(\Delta_i), \quad (24)$$

cf. (4) and (8), where the $\pi(k)$'s and the $f_k(\cdot)$'s are estimated with reasonable precision, and where $R_k(\Delta_i)$ carries information on the unknown parameters p, q, r (or β).

Choose M pixel crosses, not in the terrain, but in the computer, lying sufficiently distant from each other to make independence between their Δ_i 's a reasonable assumption. The crosses can e.g. have centres constituting a lattice of points with smallest distance equal to 5 pixel sizes. Then $\Delta_1, \dots, \Delta_M$, a collection of $5M$ spectral vectors, have simultaneous density $\prod_{i=1}^M f(\Delta_i)$, and we can employ the maximum likelihood program.

We have, assuming model (7) and using (8),

$$\begin{aligned} \frac{\partial \log f(\Delta_i)}{\partial q} &= \frac{1}{f(\Delta_i)} \sum_{k=1}^K \pi(k) f_k(\Delta_i) \{B_k(\Delta_i) - A_k(\Delta_i)\} \\ &= \sum_{k=1}^K P(k | \Delta_i) \frac{B_k(\Delta_i) - A_k(\Delta_i)}{R_k(\Delta_i)} \\ &= \sum_{k=1}^K P(k | \Delta_i) \left\{ \frac{1}{q} \beta_k(\Delta_i) - \frac{1}{p} \alpha_k(\Delta_i) \right\}, \end{aligned}$$

cf. also (4), where $p = 1 - q - r$ is used, and where

$$\begin{aligned} \alpha_k(\Delta_i) &= p A_k(\Delta_i) / R_k(\Delta_i), \\ \beta_k(\Delta_i) &= q B_k(\Delta_i) / R_k(\Delta_i), \\ \gamma_k(\Delta_i) &= r C_k(\Delta_i) / R_k(\Delta_i). \end{aligned} \tag{25}$$

Notice that these three sum to unity.

In this way we obtain maximum likelihood equations for p, q, r :

$$\begin{aligned} \frac{1}{q} \sum_{i=1}^M \sum_{k=1}^K P(k | \Delta_i) \beta_k(\Delta_i) &= \frac{1}{p} \sum_{i=1}^M \sum_{k=1}^K P(k | \Delta_i) \alpha_k(\Delta_i), \\ \frac{1}{r} \sum_{i=1}^M \sum_{k=1}^K P(k | \Delta_i) \gamma_k(\Delta_i) &= \frac{1}{p} \sum_{i=1}^M \sum_{k=1}^K P(k | \Delta_i) \alpha_k(\Delta_i). \end{aligned}$$

These can also be written

$$\begin{aligned} p &= \frac{1}{M} \sum_{i=1}^M \sum_{k=1}^K P(k | \Delta_i) \alpha_k(\Delta_i), \\ q &= \frac{1}{M} \sum_{i=1}^M \sum_{k=1}^K P(k | \Delta_i) \beta_k(\Delta_i), \\ r &= \frac{1}{M} \sum_{i=1}^M \sum_{k=1}^K P(k | \Delta_i) \gamma_k(\Delta_i). \end{aligned} \tag{26}$$

These are only implicit equations since p, q, r are involved in the r.h.s. expressions. However, solutions p^*, q^*, r^* can be obtained by using (26) at "iterating equations": start out with $(p_{(0)}, q_{(0)}, r_{(0)}) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ (say), use these values everywhere on the r.h.s. of (26), and get next generation values $(p_{(1)}, q_{(1)}, r_{(1)})$. Continue this process until a convergence criterion is met; the limit point is (p^*, q^*, r^*) .

Several comments are in order here. It was implicitly assumed above that $\hat{\pi}(k)$'s and $\hat{f}_k$'s were used, i.e. $\pi(k)$ and f_k are considered "known" and equal to their estimates. This makes sense since these are estimated with good precision from class training sets; p, q, r are for the moment the "principal unknowns". Secondly, p^*, q^*, r^* obtained above depended only on the chosen distant pixel crosses $\Delta_1, \dots, \Delta_M$. One may move this lattice to a new position and calculate $(p^*)', (q^*)', (r^*)'$ based on $\Delta'_1, \dots, \Delta'_M$, and so on. Eventually an average should be made to define the "final" estimates p^*, q^*, r^* . Thirdly, it is not obvious that the computational procedure described will converge. That it in fact will converge is a consequence of the general theory of EM-algorithms, cf. Wu (1983). General comments and other examples of "updating procedures" using unclassified vectors are in Hjort (1985b).

Finally in this subsection we show how Owen's parameter β can be estimated using the scene's unclassified vectors. The underlying model is (7) with

$$p = 1 - \beta, \quad q = .4142\beta, \quad r = .5858\beta.$$

We get (24) again with

$$R_k(\Delta_i) = (1 - \beta)A_k(\Delta_i) + .4142\beta B_k(\Delta_i) + .5858\beta C_k(\Delta_i).$$

The maximum likelihood equation for β based on $\Delta_1, \dots, \Delta_M$ becomes

$$\sum_{i=1}^M \sum_{k=1}^K P(k | \Delta_i) \frac{-A_k(\Delta_i) + .4142B_k(\Delta_i) + .5858C_k(\Delta_i)}{R_k(\Delta_i)} = 0. \quad (27)$$

The solution β^* can be arrived at by iteration.

4.3. Estimating Contextual Parameters in the 8-Neighbours Case

One may construct methods very similar to those presented in 4.2 in the case of pixel boxes, with which Section 3 was concerned. Needed are input values of either p, q, r, s, t, u or of the single Poisson field parameter α . Since the arguments are similar we just present the results.

Assume first that ground truth for a representative set of M pixel boxes is available. Let M_A be the number of these with the same class in all 9 pixels, and let $M_B, \dots, M_F$ count the number of $B, \dots, F$ patterns observed. Redefine M as $M_A + \dots + M_F$. Then M_A/M estimates $p + (1-p)w$, M_B/M estimates $q(1-w)$ and so on, where $w = \sum_{k=1}^K \pi(k)^2$. It follows that

$$\begin{aligned} \hat{q} &= \frac{M_B}{M} / (1 - \hat{w}), \dots, \hat{u} = \frac{M_F}{M} / (1 - \hat{w}), \\ \hat{p} &= 1 - \hat{q} - \hat{r} - \hat{s} - \hat{t} - \hat{u} \end{aligned} \quad (28)$$

are reasonable estimates. Next assume that the Poisson field model is in force, i.e. (19) holds. Suppose for the moment that $(M_A, \dots, M_F)$ is multinomial, which is the case if the M pixel boxes are randomly chosen from a large homogeneous scene. Then the likelihood is

$$(1 - \alpha\bar{w})^{M_A} (.2929\alpha\bar{w})^{M_B} \dots (.1188\alpha\bar{w})^{M_F}$$

where $\bar{w} = 1 - w = 1 - \sum_{k=1}^M \pi(k)^2$. Inserting the estimate $\hat{w}$ for w and focusing interest on α we find the maximum likelihood estimator

$$\hat{\alpha} = \left(1 - \frac{M_A}{M}\right) / (1 - \hat{w}) = 1 - \hat{p}. \quad (29)$$

This is a reasonable estimator also when the pixel boxes have been chosen systematically (but representatively).

p, q, r, s, t, u , and α , can also be estimated automatically, fitting the assumed stochastic model to the available, unclassified data. Δ_i of (13) has density $f(\Delta_i) = \sum_{k=1}^K \pi(k) f_k(X_i)$

$R_k(\Delta_i)$, where $R_k(\Delta_i)$ is given in (15), under model (14). Write $R_k(\Delta_i) = \prod_{j=1}^8 f_k(X_{i,j}) S_k(\Delta_i)$, where

$$S_k(\Delta_i) = pA_k(\Delta_i) + qB_k(\Delta_i) + rC_k(\Delta_i) + sD_k(\Delta_i) + tE_k(\Delta_i) + uF_k(\Delta_i).$$

Using $P(k | \Delta_i) = \pi(k) f_k(X_i) R_k(\Delta_i) / f(\Delta_i)$ we get

$$\begin{aligned} \frac{\partial \log f(\Delta_i)}{\partial q} &= \frac{1}{f(\Delta_i)} \sum_{k=1}^K \pi(k) f_k(X_i) \prod_{j=1}^8 f_k(X_{i,j}) \{B_k(\Delta_i) - A_k(\Delta_i)\} \\ &= \sum_{k=1}^K P(k | \Delta_i) \{B_k(\Delta_i) - A_k(\Delta_i)\} / S_k(\Delta_i), \end{aligned}$$

writing $p = 1 - q - r - s - t - u$, etc. Estimators $p^*, q^*, r^*, s^*, t^*, u^*$ based on $\Delta_1, \dots, \Delta_M$ can be found by iterating the maximum likelihood equations

$$\begin{aligned} p &= \frac{1}{M} \sum_{i=1}^M \sum_{k=1}^K P(k | \Delta_i) \alpha_k(\Delta_i) \\ &\vdots \\ u &= \frac{1}{M} \sum_{i=1}^M \sum_{k=1}^K P(k | \Delta_i) \phi_k(\Delta_i), \end{aligned} \tag{30}$$

where $\alpha_k(\Delta_i) = pA_k(\Delta_i)/S_k(\Delta_i), \dots, \phi_k(\Delta_i) = uF_k(\Delta_i)/S_k(\Delta_i)$. Again an average over several lattices of distant pixel boxes should be made.

The maximum likelihood equation for the single parameter α in the Poisson field model becomes

$$\sum_{i=1}^M \sum_{k=1}^K \frac{P(k | \Delta_i)}{S_k(\Delta_i)} \{-1 + .2929B_k(\Delta_i) + \dots + .1188F_k(\Delta_i)\} = 0, \tag{31}$$

cf. (20) and (27).

Acknowledgements. The present paper has been written during my stay at Stanford University with grants from the Norwegian Computing Center and the Royal Norwegian Council for Scientific and Industrial Research. The Research Council has also supported a research group at the Computing Center working with the analysis of remotely sensed data. I thank Erik Mohn and the other participants of that research group for inspiring discussions. I finally thank Paul Switzer at Stanford for encouragement and helpful comments.

References

- Haslett, J. (1983). Maximum likelihood discriminant analysis on the plane using a Markovian model of spatial context. Department of Statistics, Trinity College, Dublin.
- Hjort, N. L. (1985a). Detecting outliers. Chapter 6 in "Notes on the Theory of Statistical Symbol Recognition". Research report, Norwegian Computing Center, Oslo.
- Hjort, N. L. (1985b). Estimating parameters in neighbourhood based classifiers for remotely sensed data, using unclassified vectors. In Sæbø et al.
- Hjort, N. L. and Mohn, E. (1984). A comparison of some contextual methods in remote sensing classification. In *Proceedings of the Eighteenth International Symposium on Remote Sensing of Environment*. CNES, Paris.
- Hjort, N. L., Mohn, E., and Storvik, G. (1985). Contextual classification of remotely sensed data based on an autocorrelation model. In Sæbø et al.
- Owen, A. (1984). A neighbourhood-based classifier for LANDSAT data. *Canadian Journal of Statistics* **12**, 191-200.
- Solomon, H. (1978). *Geometric Probability*. SIAM Publications, Philadelphia.
- Switzer, P. (1965). A random set process in the plane with a Markov property. *Ann. Math. Statist.* **36**, 1859-1863.
- Sæbø, H. V., Bråten, K. H., Hjort, N. L., Llewellyn, B., and Mohn, E. (1985). "Contextual classification of remotely sensed data: statistical methods and development of a system." Project Report, Norwegian Computing Center, Oslo.
- Swain, P. H., Vardeman, S. B., and Tilton, J. C. (1981). Contextual classification of multi-spectral data. *Pattern Recognition* **13**, 429-441.
- Welch, J. R. and Salter, K. G. (1971). A context algorithm for pattern recognition and image interpretation. *IEEE Trans. on Syst. SMC-1*, 24-30.
- Wu, C. F. Jeff (1983). On the convergence properties of the EM algorithm. *Ann. Statist.* **11**, 95-103.