

Implementação de um Método Bayesiano Contextual Multi-espectral Viabilizado para Classificação de Imagens de Sensoriamento Remoto.

RONEI MARCOS DE MORAES¹
JOÃO MARQUES DE CARVALHO²
MARCOS A. G. BRASILEIRO²

¹UFPb - Universidade Federal da Paraíba
Depto. de Estatística - CCEN
Cidade Universitária, s/n.
58050 João Pessoa, Pb, Brasil
ccende02@brufpb.bitnet

²UFPb - Universidade Federal da Paraíba
Depto. de Eng. Elétrica - CCT
Av. Aprígio Veloso, s/n.
58100 Campina Grande, PB, Brasil

Abstract. Systems for contextual classification of images are known in the specialized literature for their good accuracy of results and slowness of computation. For this reason, very few attempts have been made to produce practical or commercial implementations for this class of algorithms. This paper shows the results of the implementation of a contextual classification method in the SITIM system.

1. Introdução

Resumidamente, um sistema de classificação contextual é um classificador que leva em consideração não só as informações contidas no próprio ponto, mas também as informações sobre os seus vizinhos para classificá-lo. A metodologia abaixo [Moraes (1992)] mostra em linhas gerais um classificador contextual.

Têm-se K classes possíveis para classificação, com probabilidades *a priori* p_1, p_2, \dots, p_K , com densidade da classe k dada por $f_k(x)$. Reserva-se também a possibilidade de se classificar um "pixel" de forma duvidosa. Assume-se que a perda por se estar em dúvida seja e . Para o "pixel" i , temos a sua vizinhança denotada por V_i , e o seu vetor característico por \mathcal{D}_i . Se V_i consiste em n "pixels" em adição ao "pixel" central, o vetor característico será denotado por :

$$\mathcal{D}_i = \{x_i, x_{i1}, \dots, x_{in}\} \quad (1)$$

onde as variáveis aleatórias x também podem vir a ser vetores aleatórios \vec{X} , quando no caso de imagens multiespectrais. A regra de classificação Bayesiana pode ser :

$$C_i = \begin{cases} D, & \text{se } P_i(k|\mathcal{D}_i) < 1 - e, \forall k = 1, \dots, K \\ m, & \text{se } P_i(m|\mathcal{D}_i) = \max_{k \leq K} P_i(k|\mathcal{D}_i) \geq 1 - e \end{cases} \quad (2)$$

onde $P_i(k|\mathcal{D}_i)$ é a probabilidade de C_i seja igual a k dado \mathcal{D}_i , ou seja a probabilidade *a posteriori* da classe k , dado os vetores característicos em V_i . Assim,

sem entrar em muitos detalhes, tem-se a expressão :

$$P_i(k|\mathcal{D}_i) = \frac{p_k f_k(x_i) R_k(x_{i1}, \dots, x_{in})}{P_i(\mathcal{D}_i)} \quad (3a)$$

onde:

$$R_k(x_{i1}, \dots, x_{in}) = \sum_{K_1} \dots \sum_{K_n} \mathcal{H}(x, k) \mathcal{G}(k) \quad (3b)$$

onde

$$\begin{aligned} \mathcal{H}(x, k) &= h(x_{i1}, \dots, x_{in} | x_i, k, k_1, \dots, k_n) \\ \mathcal{G}(k) &= g(k_1, \dots, k_n | k) \end{aligned} \quad (3c)$$

e onde a função h descreve a densidade de probabilidade conjunta dos vetores característicos \mathcal{D}_i , dado que os "pixels" em V_i pertencem à classe k ("pixel" central), k_1, k_2, \dots, k_n . O segundo fator, função g , é a probabilidade desta configuração em particular de classes, dado que o "pixel" central pertence à classe k . A somatória é o total das possíveis configurações. Como o denominador é independente, quando da maximização, opera-se somente com o numerador.

Como exemplos desta metodologia, podem ser citados os trabalhos de Welch e Salter (1971), Landgrebe (1980). No entanto, já se sabia que o número de vizinhos era crítico e o aumento do número de vizinhos aumentava sobremaneira o tempo de processamento. Devido a esses problemas, Swain et al

(1980), Swain et al (1981), Tilton e Swain (1981), Tilton et al (1982), Wharton (1982) e Owen (1984) propuseram soluções. Saebo et al (1985) publicam um relatório técnico sobre a implementação de métodos contextuais. Neste relatório, pode-se ver nos apêndices, os trabalhos de Hjort and Mohn (1985a), Hjort (1985a), Hjort et al (1985), Hjort and Mohn (1985b) e Hjort (1985b), sobre modelos, estimadores e ajustes das regras contextuais. Em 1986, Bezdek et al publicaram um algoritmo contextual utilizando uma abordagem nebulosa - "fuzzy". Recentemente, depois de uma certa ausência na literatura especializada, aparecem os trabalhos de Klein e Press (1989);(1990), propondo um método contextual com estrutura de correlação estatística espacial e vizinhos não classificados.

De qualquer forma, o grande problema quanto ao processamento dos métodos contextuais de classificação, ainda não foi resolvido. O fator de correção contextual - apresentado na equação (3b) - com suas somatórias, proporcionais ao total de possíveis configurações, ainda está presente. Como mencionado acima, vários trabalhos se concentraram na resolução ou tentativa de reduzir o tempo computacional gasto. Na equação (3b) o número de termos é K , onde a é o número de vizinhos. Hjort (1985a) com a sua metodologia baseada em um trabalho anterior [Owen and Switzer (1982)], reduzia o número para $8K-7$. No entanto, essas reduções ainda não eram satisfatórias para a época, principalmente pelo equipamento utilizado. Atualmente, com o rápido avanço de "hardware" e novas linguagens de programação, mais rápidas e poderosas, o problema de redução da quantidade de cálculos envolvida para aumentar a velocidade deverá perder a importância.

2. O Modelo Contextual p,q,r.

No espaço de decisão C_i a classe do "pixel" i , $C_i \in \{1, 2, \dots, K, D\}$; onde D é a reserva de se classificar um "pixel" de forma duvidosa, ou seja, não se pode classificar o "pixel" i em nenhuma das K classes disponíveis. No caso de imagens multiespectrais, o "pixel" i é na realidade um vetor \vec{X} d -dimensional composto pelas medidas espectrais dos diversos canais adquiridos pelo usuário. Os vetores \vec{X} de uma classe k , seguem uma densidade de probabilidade multivariada $f_k(\vec{X})$.

A vizinhança assumida para o "pixel" i seja composta pelos seus quatro vizinhos imediatamente ao norte, ao sul, à leste e à oeste, formando uma cruz.

A probabilidade desta configuração particular de classes em cruz, dada pela densidade g , é dada por :

$$g(a, b, c, d|k) = Pr\{C_{iN} = a, C_{iE} = b, C_{iS} = c, C_{iW} = d | C_i = k\} \quad (4)$$

Deve-se ainda computar a densidade conjunta dos vetores característicos em \mathcal{D}_i dado as classes na cruz e \vec{X} , ou seja:

$$f(\mathcal{D}_i|k, a, b, c, d) = f_k(\vec{X}_i)h(\mathcal{D}_i|k, a, b, c, d, \vec{X}_i) \quad (5)$$

Segundo a escola Bayesiana [Pereira e Viana (1982) e Maritz and Lwin (1989)], a probabilidade $P(k|\mathcal{D}_i)$ é uma probabilidade *a posteriori*. Esta probabilidade deve depender dos parâmetros k e \mathcal{D}_i e da probabilidade "inicial" para o fator k que é a probabilidade *a priori* de k ; $\pi(k)$. Então, pode-se reescrever a equação (8) como:

$$\begin{aligned} P(k|\mathcal{D}_i) &= \frac{1}{f(\mathcal{D}_i)} \pi(k) \sum_{a,b,c,d} \mathcal{G}(\cdot) f(\mathcal{D}_i|k, a, b, c, d) \\ &= \frac{1}{f(\mathcal{D}_i)} \pi(k) f(\vec{X}_i) \sum_{a,b,c,d} \mathcal{G}(\cdot) \mathcal{H}(\cdot) \end{aligned} \quad (6a)$$

onde

$$\begin{aligned} \mathcal{G}(\cdot) &= g(a, b, c, d|k) \\ \mathcal{H}(\cdot) &= h(\mathcal{D}_i|k, a, b, c, d, \vec{X}_i) \end{aligned} \quad (6b)$$

e $f(\mathcal{D}_i)$ é a densidade marginal de \mathcal{D}_i , de modo que $\sum_{i=1}^K P(k|\mathcal{D}_i) = 1$

Como pode-se notar, a implementação e o custo computacional da expressão (9), principalmente pela soma quádrupla de K classes para cada "pixel", pode tornar essa abordagem impraticável. O que se tem feito é tentar uma modelagem das probabilidades g e h , de modo a tentar reduzir e simplificar a expressão (9). Para h , vários autores [Welch and Salter (1971), Swain et al (1981) e Owen (1984)] assumiram os vetores \vec{X} como sendo condicionalmente independentes dado as classes, e também o será neste trabalho:

$$\begin{aligned} f(\mathcal{D}_i|k, a, b, c, d) &= f_k(\vec{X}_i) f_k(\vec{X}_{iN}) \\ &\quad f_k(\vec{X}_{iE}) f_k(\vec{X}_{iS}) f_k(\vec{X}_{iW}) \\ h(\mathcal{D}_i|k, a, b, c, d, \vec{X}_i) &= f_k(\vec{X}_{iN}) f_k(\vec{X}_{iE}) \\ &\quad f_k(\vec{X}_{iS}) f_k(\vec{X}_{iW}) \end{aligned} \quad (7)$$

Para g , Owen and Switzer (1982) assumiram um modelo de probabilidade geométrica, assumindo premissas descritas em Moraes (1992). Isto significa que o número de classes presentes na cruz será de no máximo duas classes. Portanto, somente para estes casos, assumiremos que a probabilidade seja maior

que zero (eventualmente poderá ser zero também para alguns destes casos). Adotando esta sistemática, o número de termos sofrerá uma redução para $8K-7$, o que é bastante significativo. A figura 1 ilustra os três tipos possíveis de cruces com probabilidade maior (ou eventualmente igual) a zero:

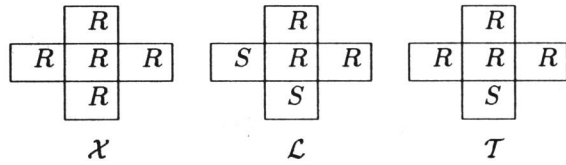


Figura 1: padrões \mathcal{X} , \mathcal{L} e \mathcal{T}

Pode-se observar que para \mathcal{X} existe somente um padrão - o padrão da figura 1 - e para \mathcal{L} e \mathcal{T} , quatro padrões possíveis pela rotação dos padrões R e S. Assim, sendo p a probabilidade do padrão \mathcal{X} , $q/4$ a probabilidade de cada padrão \mathcal{L} (q é a probabilidade do conjunto de padrões \mathcal{L}) e $r/4$ de cada padrão \mathcal{T} (raciocínio análogo a q), temos que $p+q+r=1$, pela restrição do espaço de todos os padrões possíveis. Considerando os padrões \mathcal{X} , \mathcal{L} e \mathcal{T} , tem-se que [Moraes (1992)]:

$$\begin{aligned} q(k, k, k, k \setminus k) &= p + (q + r)\pi(k) \\ g(k, k, m, m \setminus k) &= g(m, k, k, m \setminus k) = g(m, m, k, k \setminus k) \\ &= g(k, m, m, k \setminus k) = \frac{1}{4}q\pi(m) \\ g(k, k, m, k \setminus k) &= g(k, k, k, m \setminus k) = g(m, k, k, k \setminus k) \\ &= g(k, k, m, k \setminus k) = \frac{1}{4}r\pi(m) \end{aligned} \tag{8}$$

para $m, k = 1, 2, \dots, K; m \neq k$. Somente esses $8K - 7$ termos são positivos.

Esta especificação para g e a especificação (7) para h determinam um procedimento de classificação contextual. Pela especificação (7), $R_k(\mathcal{D}_i)$ pode ser reescrito como [Moraes (1992)]:

$$R_k(\mathcal{D}_i) = pA_k(\mathcal{D}_i) + qB_k(\mathcal{D}_i) + rC_k(\mathcal{D}_i) \tag{9}$$

onde :

$$\begin{aligned} A_k(\mathcal{D}_i) &= f_k(\vec{X}_{iN})f_k(\vec{X}_{iE})f_k(\vec{X}_{iS})f_k(\vec{X}_{iW}) \\ B_k(\mathcal{D}_i) &= \frac{1}{4} \left[f_k(\vec{X}_{iN})f_k(\vec{X}_{iE})b(\vec{X}_{iS}, X_{iW}) + \right. \\ &\quad + f_k(\vec{X}_{iS})f_k(\vec{X}_{iW})b(\vec{X}_{iN}, X_{iE}) + \\ &\quad + f_k(\vec{X}_{iE})f_k(\vec{X}_{iS})b(\vec{X}_{iW}, X_{iN}) + \\ &\quad \left. + f_k(\vec{X}_{iW})f_k(\vec{X}_{iN})b(\vec{X}_{iE}, X_{iS}) \right] \end{aligned}$$

e;

$$\begin{aligned} C_k(\mathcal{D}_i) &= \frac{1}{4} \left[f_k(\vec{X}_{iN})f_k(\vec{X}_{iE})f_k(\vec{X}_{iS})a(\vec{X}_{iW}) + \right. \\ &\quad + f_k(\vec{X}_{iN})f_k(\vec{X}_{iE})a(\vec{X}_{iS})f_k(\vec{X}_{iW}) + \\ &\quad + f_k(\vec{X}_{iN})a(\vec{X}_{iE})f_k(\vec{X}_{iS})f_k(\vec{X}_{iW}) + \\ &\quad \left. + a(\vec{X}_{iN})f_k(\vec{X}_{iE})f_k(\vec{X}_{iS})f_k(\vec{X}_{iW}) \right] \end{aligned} \tag{10}$$

e onde :

$$\begin{aligned} a(x) &= \sum_{m=1}^K \pi(m)f_m(x) \\ b(x, y) &= \sum_{m=1}^K \pi(m)f_m(x)f_m(y) \end{aligned} \tag{11}$$

Obviamente, o modelo ainda não está completo. Há a necessidade de se estimar os parâmetros $p, q, r, \pi(k)$ e $f_k(\vec{X})$.

2.1. Densidades das classes.

Usualmente, dispõe-se de pelo menos um conjunto de treinamento para cada classe. Então considera-se a existência de vetores observados $\vec{X}_1^{(k)}, \dots, \vec{X}_{n_k}^{(k)}$ de "pixels" dos quais são conhecidas as verdadeiras classes. Através destes conjuntos de treinamento, geralmente escolhidos em áreas homogêneas situadas seguramente dentro das classes apropriadas, serão obtidas as densidades $f_k(\vec{X})$, as quais em um caso mais geral, distribuem-se segundo normais d-dimensionais com vetores de médias $\vec{\mu}_k$ e matriz de covariâncias Σ_k para cada classe k . Pode-se estimar $\vec{\mu}_k$ e Σ_k , pelos estimadores:

$$\hat{\vec{\mu}}_k = \frac{1}{n_k} \sum_{j=1}^{n_k} \vec{X}_j^{(k)} \tag{12}$$

$$\hat{\Sigma}_k = \frac{1}{n_k - 1} (\vec{X}_k^{(k)} - \hat{\vec{\mu}}_k)(\vec{X}_k^{(k)} - \hat{\vec{\mu}}_k)' \tag{13}$$

2.2. Probabilidades a priori e parâmetros contextuais.

Da mesma forma como nas estimativas dos parâmetros das classes, as probabilidades a priori $\pi(k)$ serão estimadas de com base em amostras tomadas na própria imagem.

Seja C_{ij} a classe de um "pixel" arbitrário j na cruz de número i . Convenientemente, serão expressos os estimadores com a ajuda da função indicadora:

$$I(C_{ij}, k) = \begin{cases} 1, & \text{se } C_{ij} = k \\ 0, & \text{caso contrário} \end{cases} \tag{14}$$

onde $i = 1, \dots, n; j = 1, 2, 3, 4, 5$ correspondente aos "pixels" da cruz conforme mostra a figura 2:

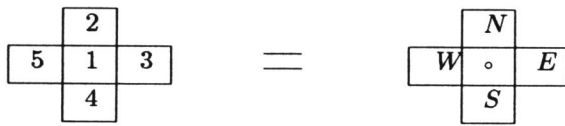


Figura 2: correspondência e localização dos "pixels" na cruz.

O estimador natural para a probabilidade *a priori* da classe k , $\pi(k)$ é :

$$\pi(k) = \frac{1}{5M} \sum_{i=1}^M \sum_{j=1}^5 \mathcal{I}(C_{ij}, k) \quad (15)$$

que nada mais é que a proporção dos "pixels" da classe k na amostra tomada.

Para a estimativa de p, q, r , das M cruces amostradas, tomam-se M cruces como as que tem cinco "pixels" de alguma classe k ; M cruces para as que tem tres "pixels" de uma classe k e dois de uma classe m qualquer; e M cruces para as que tem quatro "pixels" de uma classe k e um de uma classe m . Então M será: $M = M_X + M_C + M_T$. Estima-se a razão M_X/M por:

$$\begin{aligned} \left(\frac{M_X}{M}\right) &= \pi(k)g(k, k, k, k \setminus k) \\ &= \sum_{k=1}^K \pi(k)[p + (1 - p)\pi(k)] \\ &= p + (1 - p)w \end{aligned} \quad (16)$$

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

As estimativas para M_C/M e M_T/M são respectivamente:

$$\begin{aligned} \left(\frac{M_C}{M}\right) &= \sum_{k=1}^K \sum_{m=1, m \neq k}^K \pi(k)q\pi(m) \\ &= q(1 - w) \end{aligned} \quad (17)$$

$$\begin{aligned} \left(\frac{M_T}{M}\right) &= \sum_{k=1}^K \sum_{m=1, m \neq k}^K \pi(k)r\pi(m) \\ &= r(1 - w) \end{aligned} \quad (18)$$

Assim sendo, obtém-se os estimadores para p, q, r :

$$\begin{aligned} \hat{q} &= \frac{M_C}{1 - \hat{w}}; \\ \hat{r} &= \frac{M_T}{1 - \hat{w}}; \\ \hat{p} &= 1 - \hat{q} - \hat{r} = \frac{M_X - \hat{w}}{1 - \hat{w}} \end{aligned} \quad (19);$$

onde $\hat{w} = \sum_{k=1}^K \pi(k)^2$; é um estimador de w .

2.3. O limiar $(1 - e)$.

O limiar $(1 - e)$ contido na expressão (2), pode ser escolhido pela conveniência do usuário, podendo ser inclusive um vetor com valores distintos, ponderando diferentemente as classes em questão. Pode-se ressaltar também que para deixar o usuário mais à vontade com o sistema, a premissa que a soma dos valores deve resultar na unidade foi abolida, podendo o valor de entrada ser qualquer valor no intervalo $(0, 1]$. Isto altera a filosofia do limiar de rejeição contido na expressão (2). Neste caso, como $(1 - e)$ será escolhido pelo usuário e um valor alto pode significar uma ou mais classes sem pontos classificados em seu nome, foi utilizado um processo que pondera seu valor para que este se ajuste a cada caso, para que e seja um corte efetivo para a classe em questão.

Sendo então um corte $e' = (1 - e)$ definido pelo usuário, geometricamente ele será interpretado como uma fatia (no caso unidimensional) da distribuição a ser descartada. Por isto, se for tomada a probabilidade de $e'/2$ na Normal padrão $[0, 1]$, multiplicada ao seu desvio-padrão e subtraída sua média estimada na amostragem de áreas homogêneas, resulta em um novo ponto p_e cuja densidade equivalerá área da distribuição a ser descartada. Estatisticamente falando, tomar a densidade no ponto p_e significa descartar as caudas da distribuição normal na faixa de probabilidade $\alpha = (1 - e'/2) * 100\%$, restando o corpo central da densidade [Moraes (1992)].

No caso n-dimensional, este estratagema significa eliminar o volume proporcionado pela cauda no espaço, restando o volume do corpo central. No entanto, para isto seria necessário o cálculo de integrais de volume n-dimensionais. Para se evitar isto, considerou-se a seguinte estratégia: se for tomada cada densidade unidimensional ao invés da multivariada, o volume a ser considerado será de no mínimo $\alpha\%$, já que o volume se tornará um cubo e não mais um elipsóide. Em um caso mais geral, se houver um

limiar para cada classe, teremos $(1 - e_k)$. Neste caso, a operação será [Moraes (1992)]:

$$corte_k = f_k[\bar{\mu}_k \pm Z(e'_k) * \Sigma_{kk}^{1/2}] \quad (20)$$

onde $\bar{\mu}_k$ é o vetor de médias da classe k e Σ_{kk} é a matriz de covariâncias, tomada apenas a sua diagonal. Assim, o corte será o valor a ser excedido na densidade multivariada para que o algoritmo possa classificar o "pixel" i em alguma das classes disponíveis [Moraes (1992)].

O corte dado por (20) substituirá $(1 - e)$ na expressão (2), sendo fixado valores "default", para o caso do usuário que não quiser ponderar diferentemente cada classe, ou cujo interesse pese de forma uniforme sobre as classes em questão.

3. Otimizações dos cálculos.

Sabendo-se que os cálculos são por demais repetitivos, procurou-se minimizar o tempo gasto nestas execuções. Assim, adotou-se o cologaritmo na expressão

$$P(k \setminus \mathcal{D}_i) = \pi(k) f_k(\vec{X}_i) R_k(\mathcal{D}_i) \quad (21)$$

$$\Rightarrow \text{coln}[P(k \setminus \mathcal{D}_i)] = \text{coln}[\pi(k)] + \text{coln}[f_k(\vec{X}_i)] + \text{coln}[R_k(\mathcal{D}_i)] \quad (22)$$

O que o $\text{coln}(x)$ faz é transformar a maximização em minimização. Se a soma:

$$\text{coln}[P(k \setminus \mathcal{D}_i)] = \text{coln}[\pi(k)] + \text{coln}[f_k(\vec{X}_i)] \quad (23)$$

for maior que $\text{coln}(corte_k)$, antes do cômputo de $R_k(\mathcal{D}_i)$, ou ultrapassar uma outra classe antes deste mesmo cômputo, o cálculo para esta classe k é abandonado, buscando-se uma nova classe, se houver. Isto otimiza os cálculos, já que o maior tempo computacional gasto pela rotina será no cálculo de $R_k(\mathcal{D}_i)$.

Classificando assim, o ponto na classe k onde

$$C_i = \begin{cases} k, & \text{se } \text{coln}P(k \setminus \mathcal{D}_i) = \min_{m \leq K} P(m \setminus \mathcal{D}_i) \\ & \leq \text{coln}(corte_k) \\ D, & \text{se } \text{coln}P(k \setminus \mathcal{D}_i) > \text{coln}(corte_k), \\ & k \in [1, K] \end{cases} \quad (24)$$

o que modifica a forma da equação (2).

Os cálculos das densidades $f_k(\vec{X})$ também foram otimizados. A densidade a ser calculada segue uma distribuição normal multivariada, segundo a forma:

$$P[W_j \setminus \vec{X}] = \frac{\exp\left(-\frac{1}{2}(\vec{X} - \vec{M}_j)^t C^{-1}(\vec{X} - M)\right)}{(2\pi)^{1/2}} \quad (25)$$

sofreu transformações algébricas [Velasco et al (1978)] de modo a reduzir a distância de Mahalanobis ao quadrado:

$$r_j^2 = (\vec{X} - \vec{M}_j)^t C^{-1} (\vec{X} - \vec{M}_j) \quad (26)$$

à distância Euclidiana do ponto \vec{X} à origem ao quadrado, ou seja:

$$r^2 = (\vec{X} - \vec{M}_1)^2 + (\vec{X} - \vec{M}_2)^2 + \dots + (\vec{X} - \vec{M}_n)^2 \quad (27)$$

fazendo com que os cálculos se tornem mais rápidos, por não serem necessárias operações matriciais.

4. Implementação.

Objetivou-se implementar um sistema de classificação contextual compatível com a família SITIM desenvolvida pelo INPE e produzido pela Engespaço de São José dos Campos - S.P. O sistema foi implementado sob sistema operacional MS-DOS versão 4.01 da Microsoft e linguagem de programação C, também da Microsoft, versão 5.01. O aproveitamento das rotinas já incorporadas ao sistema SITIM, foi feito com objetivo de que o usuário final se sentisse mais a vontade com o novo sistema e também porque se criaria um sistema mais aperfeiçoado, pelo fato de suas rotinas já serem otimizadas e exaustivamente testadas. Por estes motivos, foram aproveitadas as rotinas de leitura, e as de aquisição de amostras homogêneas do MAXVER. As características desta rotina não foram alteradas, podendo o sistema classificar, como antes, até 20 classes diferentes e adquirir até 100 amostras [SITIM].

A aquisição das M amostras de cruces de "pixels" para a estimativa dos parâmetros das classes foi implementada. Esta rotina é capaz de amostrar cruces de "pixels" diretamente sobre a imagem, ter até duas classes distintas em cada cruz. O sistema deixa a cargo do usuário a definição da classe, devendo esta obrigatoriamente estar incluída entre as classes amostradas anteriormente com as rotinas MAXVER. Se acaso o usuário não tomar amostras para todas as classes contidas no arquivo MAXVER, o programa aletará ao usuário a falta de classes não amostradas.

A primeira estimativa das razões M_X/M , M_L/M e M_T/M é tomada como a própria razão das amostras, gerando assim a primeira estimativa para p , q e r . A seguir, inicia-se o processo iterativo para o refinamento da estimativa, dado pelas equações (16), (17), (18) e (19). O processo sers interrompido quando o teste iterativo

$$\frac{|\hat{p}^{(i)} - \hat{p}^{(i-1)}| + |\hat{q}^{(i)} - \hat{q}^{(i-1)}| + |\hat{r}^{(i)} - \hat{r}^{(i-1)}|}{3} < \delta \quad (28)$$

for satisfeito. Foi adotado o valor de $\delta < 0.00001$ ou 1×10^{-5} como chave de saída.

O fator de correção contextual $R_k(\mathcal{D}_i)$, dado pelas equações (9), (10) e (11) é calculado agrupando os resultados anteriores, já na forma dada por estas equações que já englobam as funções $g(o \setminus o)$ e $h(o \setminus o)$.

A forma de cálculo para a probabilidade *a posteriori*

$$P(k \setminus \mathcal{D}_i) = \pi(k) f_k(\vec{X}_i) R_k(\mathcal{D}_i) \quad (29)$$

é usada para todas as K classes. Deste cálculo, será tomado o máximo valor, e comparado com o valor de corte, calculado segundo a expressão (20), para a tomada da decisão segundo (24). É colocado também em aberto a opção de se alterar os valores de corte para cada classe - no máximo de 20 para guardar compatibilidade com o MAXVER já implementado - para adequar o sistema a diferentes formas de utilização e permitindo maior flexibilidade e iteração com o usuário. A imagem gerada ao final da classificação, pode ser visualizada no canal gráfico ou exportada para disco, utilizando-se de rotinas já disponíveis.

O módulo segue a estrutura SITIM, permitindo ao usuário trabalhar de forma independente com cada programa. A aquisição das amostras homogêneas fica por conta das rotinas do MAXVER. As amostras em cruz, estão em uma estrutura denominada MPQR. O arquivo de saída CONTEXTO.CON contém informações sobre o número total de classes, as probabilidades *a priori* $\pi(k)$ e naturalmente p, q, r . Outra rotina denominada CONTEXT contém as funções $h(o \setminus o)$, $g(o \setminus o)$ e $R_k(o)$, lendo as saídas dos arquivos MAXVER e MPQR, para calcular as referidas funções. A rotina FUNCAOE faz a ligação com o usuário com respeito ao corte, calculando os valores "default" a partir do valor de K fornecido por MPQR e permitindo alterações, se necessário for. Esta rotina gera um arquivo de saída E.CON com os valores do corte para leitura posterior.

A rotina seguinte, é o programa classificador MXVCON que se encarrega de calcular *a posteriori* e classificar cada pixel, utilizando o módulo CONTEXT como subrotina e ler a função e' do arquivo E.CON, fazendo as transformações necessárias para a sua utilização na classificação como um limiar de corte efetivo. Usa subrotinas para calcular $f_k(\vec{X})$, e as funções $a(x)$ e $b(x, y)$ dadas pela expressão (11). Ela também se encarrega de colocar de forma adequada a imagem para que as rotinas já residentes no SITIM possam se encarregar da visualização ou armazenamento da mesma. Pelo fato de trabalhar

com vizinhos e com vários planos de imagem ou bandas, a forma de leitura da imagem utiliza a técnica de "buffer" rotativo [Moraes (1992)]. No entanto, por questões de requisitos de memória, o número de bandas que o usuário poderá adquirir da imagem em uso é limitado a tres e acomodadas em planos de memória que são denominados canais. O sistema utiliza um quarto canal exclusivamente para gerar a imagem classificada.

Resumidamente, a técnica do "buffer" rotativo compõe-se de manter para cálculos um número fixo de linhas e a cada nova iteração, desprezar-se a primeira linha, rearranjando o "buffer" de forma que a nova linha a ser lida seja incorporada no final.

O arquivo de subrotinas CONTEXT contém as seguintes subrotinas: CONTEXT(), DENSID() e CLASSIFICA_IMAGEM_CONT().

A subrotina CLASSIFICA_IMAGEM_CONT() recebe de MXVCON os parâmetros de contexto.con e e.con, le a imagem através do "buffer" rotativo e classifica a imagem, produzindo a saída classificada no canal 3 da Unidade Visualizadora de Imagem - UVI. Utiliza como suas subrotinas CONTEXT() e DENSID().

A subrotina CONTEXT() recebe parâmetros de contexto.con e valores das cruces enviados por CLASSIFICA_IMAGEM_CONT(). Usa DENSID() para calcular as densidades $f_k(\vec{X})$, e no seu corpo calcular fator de correção $R_k(o)$ - equações (10) e (11) para cada caso.

A subrotina DENSID() faz o cálculo das densidades normais multivariadas, segundo a equação (25) e recebe, para realizar a tarefa, os parâmetros de CLASSIFICA_IMAGEM_CONT() devolvendo o valor numérico da densidade pedida.

O fluxograma da figura 3, mostra o sistema na sua forma final:

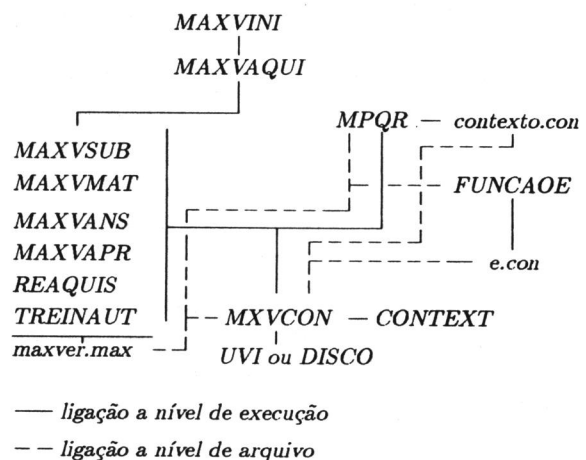


Figura 3. Fluxograma do Sistema Contextual implementado.

5. Testes e Avaliação de Desempenho.

Neste tópico, será feita uma análise dos testes do sistema contextual implementado e avaliar-se-á o seu desempenho em comparação com o sistema MAXVER que já é parte integrante do sistema SITIM, do qual o sistema contextual herdou várias de suas características.

O "hardware" utilizado é composto por um micro-computador compatível PC/AT-386-SX - com "clock" de 25MHz, memória principal de 4Mb, com co-processador aritmético 80387 e monitor DATANAV. A esta máquina, são acrescentadas quatro placas dotadas de alimentação e endereçamento independentes para trabalho específico com imagens, compondo o sistema SITIM-150.

Também foi desenvolvida uma versão para o sistema SITIM-340, cujo "hardware" é composto por um micro compatível PC/AT-286 - com "clock" de 20MHz, memória principal de 1Mb, também com co-processador aritmético 80287 e monitor NEC Multisync 3D - modificado pela inclusão de uma placa apenas, no próprio barramento do computador para o trabalho específico com imagens.

A imagem utilizada foi obtida pelo satélite LANDSAT-TM da Represa Balbina no estado do Amazonas, com localização aproximada de 1.5° de latitude Sul e 61° de longitude Oeste, pré-processada, pronta para a classificação.

O processamento se deu nas coordenadas do cursor ((198,182); (397,381)), para o sistema 150 e ((396,268); (755,579)) na versão para o sistema SITIM-340, visualizadas nas figuras 4 e 5. Os resultados comparativos estão na tabela abaixo:

Sistema	Janela	Tempo Total	Tempo por "pixel"
150	39.204	24 : 49 : 6 min	0,038seg
340	110.313	2 : 17 : 58 : 3 hor	0,075seg

Tabela 1. Quadro remissivo de desempenho.

Foi feita também uma breve avaliação comparativa entre o método de máxima verossimilhança (MAXVER), [Velasco et al (1978) e SITIM] e o método contextual (MXVCON) implementado.

Para se ter uma base melhor para tecer comentários e avaliações, a janela de classificação foi a mesma, gerando as imagens das figuras 5a e 5b para ambos os sistemas. O desempenho comparativo entre os sistemas MXVCON e MAXVER são analisados no quadro a seguir (em km^2):

Classe	Área classificada por cada sistema			
	MAXVER		MXVCON	
	SITIM 150	SITIM 340	SITIM 150	SITIM 340
Floresta	63,3000	56,4678	67,5837	60,8616
Rio	30,800	31,5756	35,7507	36,2232
Solo	0,500	0,5292	1,2231	1,3779
Não Classificado	13,400	12,5154	1,2933	1,3779

Tabela 2. Áreas classificadas pelos sistemas MAXVER e Contextual.

Pode-se notar pela tabela 2 a existência de uma superioridade do método contextual sobre o método de máxima verossimilhança. Numericamente, pode-se notar a área não classificada por cada método, a proporção não classificada chega a ser 10,36 vezes menor no método contextual. Observando as imagens geradas, principalmente na área de Solo, nota-se o rendimento bastante superior do método contextual. Nota-se que na classe Rio, há um trecho do rio que não foi reconhecido pelo MAXVER.

Uma rápida comparação entre os desempenhos dos métodos em cada sistema SITIM, pode ser vista na tabela 3:

Sistema	Tempo de Processamento	
	MAXVER	MXVCON
SITIM340	22:10 seg	24:49:6 min
SITIM150	2:39:3 min	2:17:58:3 hor

Tabela 3. Tempos de processamento por sistemas dos programas MAXVER e MXVCON

Nota-se que pelo método contextual, o tempo gasto foi de 2:17:58:3 horas. Este tempo muito superior, no entanto, tem uma explicação; observa-se que a área classificada é aproximadamente a mesma nos dois casos e há uma diferença significativa do número de pontos calculados. A diferença é a resolução do monitor. O monitor NEC-3D tem uma resolução muito superior a do DATANAV e visto que ambos os classificadores lêem a imagem da tela, explica-se por que o tempo de classificação foi tão diferente.

Pôde-se observar também na tabela 2 que houve uma alteração das áreas classificadas a nível geral. Isto se deve ao fato da resolução maior do monitor de vídeo NEC-3D, melhorando a escala, permitindo maior precisão na estimativa de área. Possivelmente, também devido às diferenças entre os monitores, a nível visual, pôde-se observar uma melhora na precisão dos dois algoritmos no sistema SITIM 340, o

que pode ser notado verificando-se as figuras 4 e 5.

6. Conclusões Finais.

Apesar da lentidão inerente ao método contextual implantado, sua utilização é desejável sempre que se necessita de uma certeza maior de classificação. Esta validade é ainda maior em termos de tempo de processamento se o tipo de monitor utilizado não for de resolução muito alta. Por um outro fator, não inerente ao método, mas acoplado a ele, o usuário tem a facilidade neste sistema de definir os limiares conforme sua vontade, exercendo uma maior supervisão sobre o trabalho do algoritmo para uma ou mais classes específicas de interesse.

Pelas simulações em diversos tipos de máquina, pode-se concluir que as otimizações na programação do algoritmo surtiram efeito e a questão da morosidade nos cálculos da função de correção contextual $R_k(\mathcal{D}_i)$ é somente uma questão de avanço de "hardware" para suas proporções serem reduzidas.

Agradecimentos.

Gostaríamos de agradecer a preciosa colaboração do D.P.I. do INPE, principalmente nas pessoas dos professores Nelson D. A. Mascarenhas e Gilberto Câmara e ao CEPAGRI da UNICAMP, sem a qual seria impossibilitada a realização deste trabalho.

Referências.

- R. M. de Moraes, Implementação de um sistema contextual de classificação de imagens orbitais compatível com o sistema SITIM. Dissertação de Mestrado, UFPb, 1992, 100 pp.
- J.R. Welch and K.G. Salter, A context algorithm for pattern recognition and image interpretation. *IEEE Transactions on Syst., Man and Cybernetics*, vol.SMC-1, n.1, January, 1971, pp 24-30.
- D.A. Landgrebe, The development of a spectral-spatial classifier for Earth observational data. *Pattern Recognition*, vol.12, 1980, pp 165-175.
- P.H. Swain, H.J. Siegel and B. W. Smith, Contextual Classification of Multispectral Remote Sensing Data using a Multiprocessor System. *IEEE Transactions on Geoscience and Remote Sensing*, vol.GE-18, n.2, April, 1980, pp 197-203.
- P.H. Swain, S.B. Vanderman and J.C. Tilton, Contextual Classification of Multispectral Image Data. *Pattern Recognition*, vol.13, n.6, 1981, pp 429-441.
- J.C. Tilton, and P.H. Swain, Incorporating Spatial Context into Spatial Classification of Multidimensional Image Data. Technical Report of LARS - Purdue University, LARS - 072981, August, 1981, 92 pp.
- J.C. Tilton, S.B. Vanderman and P.H. Swain, Estimation of context for statistical classification of multispectral image data. *IEEE Transactions on Geoscience and Remote Sensing*, vol.GE-20, n.4, October, 1982, pp 445-452.
- S.W. Whorton, A contextual classification method for recognizing land use patterns in high resolution remotely sensed data. *Pattern Recognition*, vol.15, n.4, 1982, pp 317-324.
- A. Owen, A neighbourhood-based classifier for LANDSAT Data. *The Canadian Journal of Statistics*, vol.12, n.3, 1984, pp 191-200.
- H.V. Saebo, K. Braten, N.L. Hjort, B. Llewellyn, and E. Mohn, Contextual Classification of Remotely Sensed Data Statistical Methods and Development of a System - Report n.768 - Norwegian Computing Center - Oslo - 1985.
- N.L. Hjort and E. Mohn, A comparison of some contextual methods in remote sensing classification. in Saebo, H.V. et al; Contextual Classification of Remotely Sensed Data: Statistical Methods and Development of a System; Technical Report of Norwegian Computing Center, n.768, 1985, Oslo, Enclosure 2
- N.L. Hjort, Neighbourhood Based Classification of Remotely Sensed Data based on Geometric Probability Models. in Saebo, H.V. et al; Contextual Classification of Remotely Sensed Data: Statistical Methods and Development of a System; Technical Report of Norwegian Computing Center, n.768, 1985, Oslo, Enclosure 4.
- N.L. Hjort, E. Mohn and G. Storvik, Contextual Classification of Remotely Sensed Data, Based on an Autocorrelation Model. in Saebo H.V. et al; Contextual Classification of Remotely Sensed Data: Statistical Methods and Development of a System; Technical Report of Norwegian Computing Center, n.768, 1985, Oslo, Enclosure 5.
- N.L. Hjort and E. Mohn, Notes on adjustments to contextual classification rules when some neighbours are missing. in Saebo H.V. et al; Contextual Classification of Remotely Sensed Data: Statistical Methods and Development of a System; Technical Report of Norwegian Computing Center, n.768, 1985, Oslo, Enclosure 6.
- N.L. Hjort, Estimating parameters in neighbourhood based classifiers for remotely sensed data, using unclassified vectors. in Saebo H.V. et al; Contextual Classification of Remotely Sensed Data: Statistical Methods and Development of a System; Technical Report of Norwegian Computing Center, n.768, 1985, Oslo, Enclosure 7.
- J.C. Bezdek, S.K. Chuah and D. Leep, Generalized k-Nearest Neighbor Rules. *Fuzzy Sets and Syst*, v.18, n.2, 1986, pg 237-256.
- R. Klein and S.J. Press, Contextual Bayesian Classi-

fication of Remotely Sensed Data. Commun. Statist. - Theory and Methods, vol.18, n.9, 1989, pp 3177-3202.

R. Klein and S.J. Press, Contextual Bayesian Classification of Remotely Sensed Data when Training Data is a part of the Scene. Revista Brasileira de Probabilidade e Estatística, vol.4, 1990, pp 43-67.

A. Owen and P. Switzer, A neighbourhood based classifier for LANDSAT Data : Technical Report, Dep. of Statistics, Stanford University, 1982.

C.A.B. Pereira e M.A.G. Viana, Elementos de Inferência Bayesiana. V Simpósio Brasileiro de Probabilidade e Estatística, São Paulo, 1982.

J.S. Maritz and T. Lwin, Empirical Bayes Methods. Chapman and Hall Ltd, Great Britain, 1989.

P. Switzer, A random set process in the plane with a Markov property. Ann. Math. Statist., vol.36, 1965, pp 1859-1863.

SITIM - Manual do Usuário

F.D.R. Velasco, L.O.C. Prado e R.C.M. Souza, Sistema MAXVER: Manual do usuário. Publicação interna INPE - 1315 - NTI/110, julho 1978.

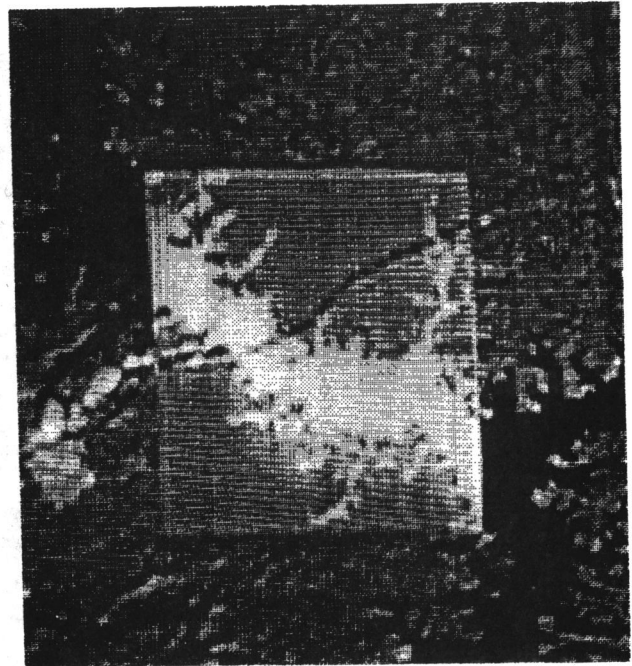


Figura 4.b. Classificação MXVCON no SITIM 150.

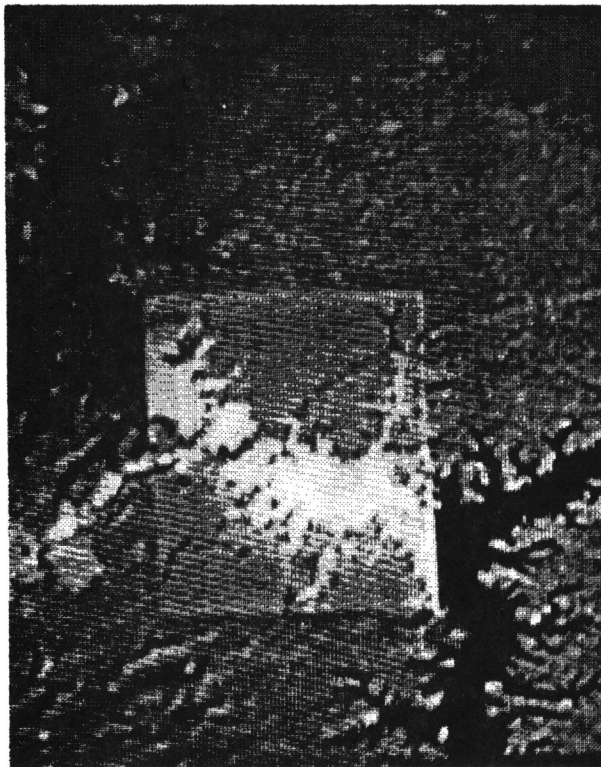


Figura 4.a. Classificação MAXVER no SITIM 150.

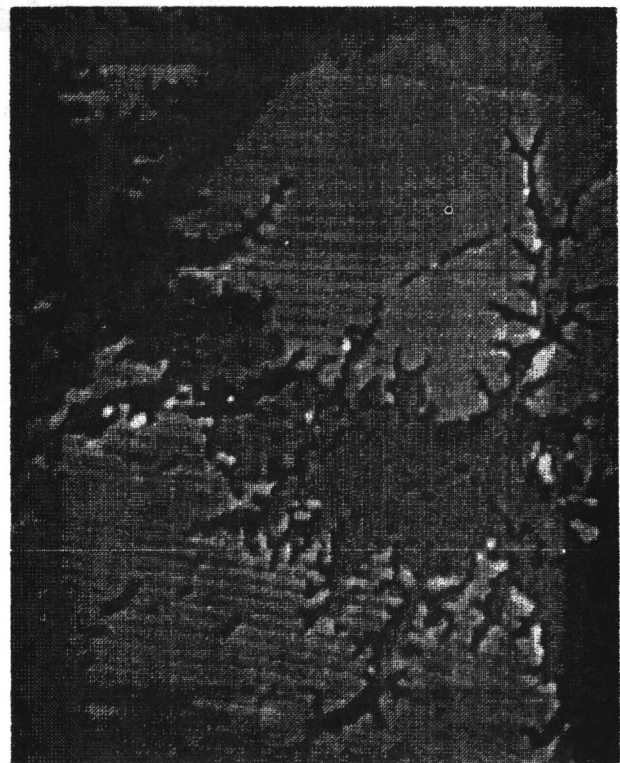


Figura 5.a. Classificação MAXVER no SITIM 340.

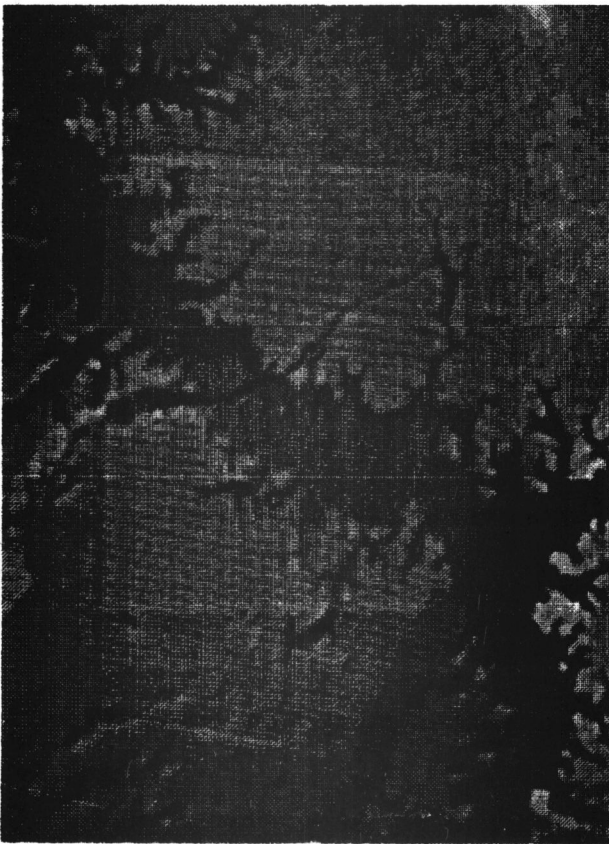


Figura 5.b. Classificação MXVCON no SITIM 340.