

CLASSIFICAÇÃO POR MÁXIMA VEROSSIMILHANÇA NUM AMBIENTE
MACIÇAMENTE PARALELO *

CELSO LUIZ MENDES
Instituto de Pesquisas Espaciais
Ministério da Ciência e Tecnologia
Caixa Postal 515, 12201 - São José dos Campos, SP, Brasil

RESUMO

Apresenta-se inicialmente a caracterização teórica do problema de classificação em imagens pelo método da máxima verossimilhança, segundo critérios estatísticos. Em seguida é feita uma análise de opções de algoritmos para implementação deste procedimento. Propõe-se, como alternativa à implementação numa máquina convencional, o desenvolvimento de um sistema paralelo para executar a classificação de forma eficiente. Após a descrição de um processador paralelo já disponível (Gapp), apresenta-se uma arquitetura maciçamente paralela, composta de uma matriz com 48x48 processadores controlados por um microcomputador de propósito geral, em desenvolvimento no INPE/DPI. Ao final comenta-se aspectos ligados ao desempenho esperado de tal sistema.

ABSTRACT

The work begins with an introduction of the theoretical basis for the problem of image classification by the maximum likelihood method, according to statistical properties. It follows with some options in terms of algorithms for implementing that operation. It is suggested, as opposed to the conventional serial approach, the development of a parallel system to carry out that procedure in an efficient way. After describing an available parallel processor (Gapp), a massively parallel architecture currently being developed at INPE/DPI is presented, with an array of 48x48 processing elements and the main control in a general purpose microcomputer. The work is concluded with some comments related to the expected speed of this system.

* - Trabalho financiado pela Sid-Informática, Projeto Extra e pela Fapesp, Projeto "Computação:87/1786-1"

1 - INTRODUÇÃO

O procedimento de classificação em imagens é frequentemente utilizado no tratamento de imagens orbitais obtidas por sensores multiespectrais. Este procedimento consiste basicamente em dividir os vários pontos da imagem em grupos ou classes, de forma que os pontos de uma dada classe tenham características espectrais similares. O critério de similaridade a ser adotado irá definir o tipo de classificador a ser empregado.

Os vários métodos de classificação podem ser, de uma forma geral, separados em dois grandes grupos: aqueles que utilizam critérios geométricos e aqueles baseados em critérios estatísticos. No primeiro grupo cada ponto da imagem é classificado segundo sua distância aos demais pontos no espaço de atributos. Apesar de eficientes sob o ponto de vista computacional, tais métodos produzem resultados não muito precisos, com erros tanto de excesso como de omissão. Por outro lado, os métodos estatísticos, apesar de mais custosos computacionalmente, produzem resultados mais confiáveis.

O método estatístico mais utilizado no tratamento digital de imagens é o da máxima verossimilhança, no qual cada classe é modelada segundo uma distribuição gaussiana. O critério de classificação utilizado é a regra de decisão Bayesiana. Tal método tem sido implementado em diversos sistemas de processamento de imagens, geralmente baseados em um processador convencional, com modo de processamento estritamente serial. O principal objetivo deste trabalho é analisar como tal método poderia ser implementado numa estrutura paralela, com arquitetura razoavelmente distinta das máquinas convencionais. São mostradas as limitações às quais uma implementação a curto prazo estaria restrita, sendo ainda apresentadas possíveis extensões futuras.

2 - MÉTODO DA MÁXIMA VEROSSIMILHANÇA

Neste método cada classe W_k é modelada por uma função densidade de probabilidade $p(X|W_k)$ do tipo gaussiana. Um certo ponto da imagem com valor espectral $X=X_0$ deve ser associado à classe "mais provável", ou seja, a classe que maximize $P(W_k|X_0)$. Pela regra de Bayes, tem-se que

$$P(W_k|X) = \frac{p(X|W_k) \cdot P(W_k)}{p(X)} \quad (I)$$

Como o valor $p(X)$ não depende das classes, basta maximizar o numerador de (I). Supondo-se que a priori as classes sejam igualmente prováveis, isto é, $P(W_1)=P(W_2)=\dots=P(W_j)$, o único termo a ser maximizado é $p(X|W_k)$, que no modelo gaussiano tem a seguinte expressão:

$$\frac{1}{(2\pi)^{n/2} |C_k|^{1/2}} \cdot \exp\left\{-\frac{1}{2} (X-M_k)^T C_k^{-1} (X-M_k)\right\}$$

onde:

- n é o número de atributos (bandas) da imagem
- M_k é o vetor-média da classe W_k
- C_k é a matriz de covariância da classe W_k

Pode-se mostrar que a expressão acima é maximizada quando for minimizado o termo

$$\ln |C_k| + R_k \quad (II)$$

onde $R_k = (X-M_k)^T C_k^{-1} (X-M_k)$ é a distância de Mahalanobis do ponto X à média M_k .

Logo, para classificar cada ponto X da imagem deve-se procurar a classe W_k que minimize a expressão (II). É comum, na prática, fixar-se um limiar L tal que, se para todas as classes W_k , ocorrer que

$$\ln |C_k| + R_k \leq L$$

então o ponto é dito não-classificado, ou rejeitado, podendo ser associado a uma classe especial W_0 .

Os valores M_k e C_k de cada uma das classes são obtidos numa etapa preliminar, denominada fase de treinamento, na qual o usuário indica ao sistema pontos na imagem representativos de cada classe (amostras). Supondo haverem G amostras numa certa classe W_k , tem-se:

$$M_k = \frac{1}{G} \sum_{i=1}^G X_i$$

$$C_k = \frac{1}{G} \sum_{i=1}^G (X_i - M_k) \cdot (X_i - M_k)^T$$

3 - FASE DE CLASSIFICAÇÃO

Conforme visto anteriormente, o processo de classificação do ponto X consiste em procurar pela classe Wk que minimize a expressão (II). Como o valor $\ln|Ck|$ é uma constante para cada classe, basta calcular o termo contendo Rk, e depois somá-lo a $\ln|Ck|$, que já pode ter sido calculado previamente. Para o cálculo do termo com Rk, há duas possibilidades:

Opção 1 - A primeira forma de cálculo do termo Rk vem de sua própria definição, e pode ser dada por :

$$Rk = \sum_{i=1}^n \sum_{j=1}^n cij \cdot (xi-mi) \cdot (xj-mj)$$

onde:

n é o número de atributos (bandas) da imagem

xi é o nível de cinza do ponto X relativo ao i-ésimo atributo (i-ésima coordenada de X)

mi é a i-ésima coordenada do vetor-média Mk

* cij é o elemento (i,j) de Ck⁻¹

Opção 2 : Através de uma transformação de coordenadas no espaço de atributos, pode-se fazer com que a nova matriz de covariância seja igual à matriz identidade, o que irá reduzir a distância de Mahalanobis à distância euclidiana. Esta transformação (Velasco,1978) é tal que:

$$X' = X - Mk, \quad X'' = Tk \cdot X'$$

Com esta transformação, Rk passa a ser dado por

$$Rk = x''_1^2 + x''_2^2 + \dots + x''_n^2$$

É possível ainda encontrar uma transformação tal que Tk seja triangular inferior, fazendo

$$Ck = Lk \cdot Lk^t, \quad Tk = Lk^{-1}$$

Desta forma, as coordenadas do ponto transformado passam a ser

$$x''_1 = t_{11} \cdot (x_1 - m_1)$$

$$x''_2 = t_{21} \cdot (x_1 - m_1) + t_{22} \cdot (x_2 - m_2)$$

$$\dots$$

$$x''_n = t_{n1} \cdot (x_1 - m_1) + \dots + t_{nn} \cdot (x_n - m_n)$$

O valor Rk pode ser dado ainda pela expressão geral

$$Rk = \sum_{i=1}^n \left(\sum_{j=1}^n t_{ij} \cdot (x_i - m_i) \right)^2$$

4 - IMPLEMENTAÇÃO SERIAL

Para implementação num processador com arquitetura seqüencial, a opção 2 mostrada acima torna-se mais interessante, já que, em geral, não é necessário calcular todas as coordenadas de X''. À medida que vai sendo acumulada a soma dos quadrados de xi'', pode-se comparar este total com o valor de limiar escolhido ou com o valor de Rk mínimo obtido nas classes anteriores. Quando a soma ultrapassa um destes valores, a classe corrente pode ser descartada. Desta forma, evita-se ter que calcular todas as coordenadas, bem como tem-se a vantagem que as coordenadas descartadas são as que envolvem maior número de termos.

Esta é a forma pela qual o algoritmo foi implementado no Laboratório de Tratamento de Imagens Digitais do Inpe (sistema I-100) e também no sistema Sitim, baseado em microcomputador.

5 - IMPLEMENTAÇÃO PARALELA

Encontra-se em desenvolvimento no Departamento de Processamento de Imagens (DPI) do Inpe um sistema paralelo destinado a aplicações gerais em processamento de imagens. Tal sistema é baseado num processador paralelo disponível comercialmente, denominado Gapp ("Geometric Arithmetic Parallel Processor"), fabricado pela NCR (EUA). Cada dispositivo contém 72 processadores do tipo serial por bit, com ciclo de operação de 10 MHz, o que implica na execução de uma instrução a cada 100 ns (Davis,1984). A partir da união de 32 destes dispositivos, pode-se formar uma matriz de 48x48 processadores básicos operando em paralelo no modo SIMD (Flynn,1972), isto é, todos os processadores executam sincronamente a mesma instrução irradiada por uma unidade central de controle, cada um deles sobre seus dados específicos.

A forma mais natural de processamento num sistema deste tipo é cada processador básico estar associado a um pixel da imagem. Desta maneira a imagem original deve ser di-

vidida em vários blocos, cada um com tamanho 48x48, os quais são processados um após o outro. Numa imagem de tamanho 512x512 haveria um total de 121 sub-imagens a serem tratadas.

No modo de operação citado, cada processador pode armazenar em sua memória local os diversos valores de nível de cinza de um pixel em suas várias bandas. Os parâmetros das classes, que são determinados na fase de treinamento, podem ser transmitidos globalmente para todos os processadores. Devido aos processadores serem do tipo serial por bit, cada operação a ser realizada levará um tempo determinado pelo número de bits dos operandos, ou seja, cada operação demora o mínimo tempo necessário para a precisão adotada. É possível mostrar que neste sistema as operações aritméticas básicas envolvendo números inteiros têm as durações mostradas a seguir, conforme os operandos sejam vetores (nas memórias locais) ou escalares (dados globais):

(A1) SOMA : vetor + vetor
dois termos na memória local, p e q bits (p) = q), resultado com p+1 bits na memória local:
duração = 2p+q+1 passos

(A2) SOMA : vetor + escalar
um dos termos na memória local, outro transmitido globalmente, p e q bits (p) = q), resultado com p+1 bits na memória local:
duração = 2p+1 passos

(B1) MULTIPLICAÇÃO : vetor x vetor
dois termos na memória local, p e q bits (p) = q), resultado com p+q bits na memória local:
duração = 6pq+2q-3p-1 passos

(B2) MULTIPLICAÇÃO : vetor x escalar
um dos termos na memória local (p bits) e outro transmitido globalmente (q bits), resultado com p+q bits na memória local:
duração = 5pq+2p-3q-1 passos

A partir destes dados, supondo-se imagens com n bandas e considerando-se o tempo de subtração igual ao da soma, é possível analisar as duas opções de implementação do algoritmo de classificação já mostradas, em termos de tempo de processamento:

Opção 1 :

(i) n subtrações :
xi-mi
vetor-escalar = vetor

(ii) (n*n+n)/2 produtos:
(xi-mi)(xj-mj)
vetor*vetor = vetor

(iii) (n*n+n)/2 produtos:
c*(xi-mi)(xj-mj)
escalar*vetor = vetor

(iv) ((n*n+n)/2)-1 somas:
Rk
vetor+vetor = vetor

Opção 2 :

(i) n subtrações:
xi-mi
vetor-escalar = vetor

(ii) (n*n+n)/2 produtos:
tij(xi-mi)
vetor*escalar = vetor

(iii) (n*n-n)/2 somas:
 \sum tij(xi-mi)
vetor+vetor = vetor

(iv) n produtos:
(\sum)*(\sum)
vetor*vetor = vetor

(v) n-1 somas:
Rk
vetor+vetor = vetor

Nos quadros a seguir estão mostrados os números de passos resultantes para as duas opções de cálculo de Rk, em função do número de bandas (1 a 4) e também do número de bits/pixel/banda (7 ou 8), que são os casos mais comuns na prática.

Opção 1 :

	7 bits	8 bits
n=1	1037	1224
n=2	3244	3815
n=3	6550	7696
n=4	10956	12868

Opção 2 :

	7 bits	8 bits
n=1	1660	2086
n=2	3957	4893
n=3	6817	8335
n=4	10011	12186

Pelos resultados mostrados, pode-se concluir que :

- (a) o número de bits/pixel (7 ou 8) é indiferente para a escolha de uma certa opção.
- (b) havendo até 3 bandas, a opção 1 é mais rápida, caso contrário (4 bandas) a opção 2 é a mais conveniente.

Este resultado deixa claro que, no caso de poucas bandas, a forma de implementação que era utilizada no caso seqüencial não é a mais conveniente para o sistema paralelo. Deve ser dito ainda que numa implementação paralela da opção 2, é pouco provável que se possa evitar o cálculo das últimas coordenadas de X", como no caso seqüencial; para que isto ocorra, é necessário que a classe seja descartada em todos os 48x48 pontos da sub-imagem que está sendo processada. Se a opção 2 for adotada, o tempo de processamento passa a ser função de cada imagem, enquanto que pela opção 1 o tempo é fixo e dado pelo quadro mostrado anteriormente.

Outra análise que se torna interessante é a de ocupação da memória interna de cada processador básico, dada a reduzida capacidade desta memória (128 bits). Seguindo as etapas mostradas acima, supondo-se imagens de r bits/pixel/banda, tem-se as seguintes taxas de ocupação da memória interna:

Opção 1 :

- (i) $n.(r+1)$ bits
- (ii) $e(iii) 3.r+2$ bits
- (iv) $3.r+2+(n*n+n)/2$ bits

Opção 2 :

- (i) $n.(r+1)$ bits
- (ii) $2.r+1$ bits
- (iii) $2.r+1+(n*n-n)/2$ bits
- (iv) $4.r+2+n*n+n$ bits
- (v) $4*r+2+n*n-1$ bits

$$\text{Total}(1) = n*n/2 + (r+3/2)*n + 6r + 3$$

$$\text{Total}(2) = n*n*5/2 + (r-1/2)*n + 12r + 5$$

$$\text{Total}(2) - \text{Total}(1) = 2*(n*n - n + 3r + 1)$$

Como esta diferença é sempre positiva para $r > 0$, pode-se concluir que a opção 2 sempre ocupa mais memória que a opção 1, e portanto a opção 1 é a mais conveniente sob o aspecto de ocupação de memória.

É possível ainda estabelecer certos limites para o uso da memória interna pela opção 1, considerando-se que deve ser reservado espaço permanente para:

- (a) pixel original : $n.r$ bits
- (b) máximo obtido para $\log|C_k| + R_k$: $3r+1+(n*n+n)/2$ bits
- (c) índice k deste máximo: $\log k$ bits
- (d) armazen.temporário: 1 bit

$$\text{Total(perm.)} : rn+3r+(n*n+n)/2+\log k+2$$

Este total permanente, somado ao que é alocado durante o cálculo de cada R_k , não pode ultrapassar 128. Daí tem-se :

$$\text{Total}(1) + \text{Total(perm.)} (= 128)$$

Desta desigualdade obtêm-se os seguintes valores máximos:

- $r=8$: $n(\text{máx.})=2$, $0 (= \log k (= 11)$
- $r=7$: $n(\text{máx.})=3$, $0 (= \log k (= 3)$
- $r=6$: $n(\text{máx.})=3$, $0 (= \log k (= 18)$
- $r=5$: $n(\text{máx.})=4$, $0 (= \log k (= 14)$

Estes resultados representam as seguintes restrições para o procedimento de classificação neste sistema pelo método apresentado:

- 2 bandas : até 8 bits/pixel/banda
- 3 bandas : até 6 bits/pixel/banda ou até 7 bits/pixel/banda, até 8 classes
- 4 bandas : até 5 bits/pixel/banda

6 - ESTIMATIVA DE DESEMPENHO

A implementação completa do algoritmo de classificação em cada sub-imagem 48x48 deve seguir os seguintes passos:

- (i) calcula $\log|C_1| + R_1 * R_1$
- (ii) compara com limiar ("classe 0")
- (iii) guarda o menor valor e seu índice
- (iv) calcula $\log|C_2| + R_2 * R_2$
- (v) compara com resultado acumulado
- (vi) guarda o menor e seu índice
- ...
- Repete (iv)(v)(vi) para as demais classes

Considerando a implementação através da opção 1 já mostrada, e assumindo que o passo (ii) acima tem uma duração igual a

$$15*(\text{num.bits}-\text{resultado})$$

e o passo (iii) tem duração igual a

$$8*(\text{num.bits}-\text{res.}) + 7*(\text{num.bits}-\text{índ.})$$

a duração dos passos (i)(ii)(iii) é composta pelos seguintes termos :

- (i) quadro acima + soma($\log|C_k|$)
- (ii) $15 * (\text{num.bits}-\text{resultado})$
- (iii) $8*(\dots) + 7*(\dots)$

Supondo uma imagem com 7 bits/pixel/banda , 3 bandas e até 7

classes, isto é, $r=7$, $n=3$ e $k=7$, tem-se:

- (i) $6550 + 57 = 6607$
- (ii) $15 \cdot 28 = 420$
- (iii) $8 \cdot 28 + 7 \cdot 3 = 245$

Total/classe = 7272 passos

7 classes --) 50904 passos

Num sistema operando a 10 MHz, isto equivale a um tempo total de 5,0904 ms, referente ao processamento efetivo. O tempo de entrada/saída de dados corresponde à entrada de $3 \cdot 7 = 21$ planos de bits com os pixels originais e à saída de 3 planos de bits contendo as classes obtidas, isto é, a movimentação de 24 planos de bits pelo sistema. Cada plano de bits pode ser movimentado em 0,0048 ms, perfazendo assim um tempo de entrada/saída de 0,1152 ms. O tempo total será portanto de 5,2056 ms por sub-imagem. Como há 121 sub-imagens, o tempo final será de 629,9 ms, ou seja, 0,63 s para classificar uma imagem 512×512 .

7 - CONCLUSÕES

Foi mostrado que a implementação do algoritmo de classificação por máxima verossimilhança num ambiente paralelo pode ser feita, em certos casos, seguindo um procedimento diferente daquele empregado no caso estritamente seqüencial, de acordo com os vários parâmetros envolvidos (número de bits, de bandas e de classes). Embora o sistema paralelo apresentado imponha fortes restrições a tais parâmetros, basicamente devido ao tamanho da memória local dos processadores, o desempenho estimado é significativamente alto para justificar sua utilização quando cabível. Isto indica que, tão logo esteja disponível um novo processador sem esta limitação de memória, assim como um novo dispositivo contendo um maior número de processadores, a implementação paralela será a mais vantajosa alternativa existente.

8 - REFERÊNCIAS BIBLIOGRÁFICAS

(1) Flynn, M.J. - "Computer Organizations and their Effectiveness", IEEE Trans. on Comp., C-21, No. 9, Sept. 1972, pp 948-960.

(2) Davis, R. & Thomas, D. - "Systolic array chip matches the pace of high-

speed processing", Electronic Design, October 1984.

(3) Velasco, F.R.D. et alii - "Sistema Maxver: Manual do Usuário", Relatório Inpe-1315, Julho 1978.