

Análise Comparativa de Técnicas de Classificação Automática de Imagens de Satélite.

LEONARDO VIDAL BATISTA¹
WEILER ALVES FINAMORE²

¹IBM Brasil/Centro Científico Rio
Av. Pres. Vargas, 824/844, 22o andar
20.071-001 Rio de Janeiro, RJ

²PUC-Rio/CETUC
Rua Marquês de São Vicente, 225
22.451-040 Rio de Janeiro, RJ

Abstract. In this paper we compare several satellite images classifiers, and variations of two of these techniques are proposed. To analyze the accuracy of the classifiers, synthetic images with known classes were used, and a measure was introduced to directly compare the classifications of real images.

1 Introdução

Na classificação por computador de imagens de satélite é comum obter uma precisão superior a 80%. Várias técnicas têm sido propostas com o objetivo de elevar este índice sem aumentar demasiadamente o custo computacional. Um aspecto importante, abordado no presente trabalho, é a comparação do desempenho de diferentes classificadores de imagens de satélite.

Dentre os classificadores não-paramétricos, que se caracterizam por não necessitarem de qualquer conhecimento prévio sobre as classes, considerou-se unicamente o método denominado K-Médias Adaptativo-Hierárquico (KMAH).

Para os classificadores paramétricos adotou-se o modelo gaussiano, o que implica a necessidade de treinamento para a estimação dos vetores médias e matrizes covariâncias, parâmetros das funções densidade de probabilidade condicionais. Optou-se aqui pelo treinamento automático, realizado a partir de uma pré-classificação pelo KMAH. Foram incluídos o classificador de máxima verossimilhança (MV); o de máxima probabilidade *a posteriori* (MAP), com probabilidades a priori estimadas iterativamente, de acordo com a classificação no passo anterior; o MAP adaptativo (MAPA), proposto por Zyl-Burnette [Zyl-Burnette (1992)]; e o classificador contextual (CX) apresentado por Tilton-Swain [Tilton-Swain (1981)].

Neste trabalho foram também propostas duas técnicas de classificação: o MAP por sub-imagens e o contextual por sub-imagens. Estas abordagens são variações do MAP adaptativo e do contextual, respectivamente. O MAP por sub-imagens se propõe

a reduzir o tempo de processamento do MAP adaptativo, conservando, no entanto, suas características básicas. O contextual por sub-imagens, por sua vez, procura explorar mais profundamente as características locais da imagem, reduzindo com isto a tendência de uniformização apresentada em alguns casos pelo contextual.

Uma das dificuldades na avaliação do desempenho dos métodos de classificação é que nem sempre se tem disponível para comparação a classificação correta da imagem. Para contornar este problema foram utilizadas como referências imagens sintéticas com classes conhecidas, geradas de forma a apresentarem características semelhantes às de imagens reais. Foi introduzida a medida de *discrepância*, para que os resultados obtidos pelos diversos classificadores sobre imagens reais pudessem ser comparados entre si.

2 Descrição dos Classificadores Usados

2.1 O K-médias Adaptativo-Hierárquico.

O método conhecido por K-médias [Duda - Hart (1973)] classifica cada pixel da imagem em exame atribuindo-o a uma das médias de um conjunto inicial M de K médias. A seguir, este conjunto é atualizado usando a divisão por classes que acaba de ser realizada, e o processo é repetido até que um critério de parada se verifique.

Um problema imediato que se coloca é a determinação das médias iniciais. Uma escolha infeliz do ponto de partida pode tornar a convergência mais demorada ou até mesmo levar a uma classificação

completamente errônea.

O K-médias Adaptativo-Hierárquico (KMAH) difere do K-médias basicamente na forma de obtenção de M . Para descrever o KMAH, considere o conjunto de N pixels da imagem, $\{x_i, i = 1, \dots, N\}$, um limiar de decisão, L , e um conjunto de médias $M' = \{m_1, m_2, \dots, m_{K'}\}$, com K' maior ou igual ao número de classes K em que a imagem deverá ser classificada.

A primeira etapa do KMAH consiste em encontrar o conjunto M' com o seguinte procedimento: faz-se $m_1 = x_1$ e o conjunto M' contém então apenas um elemento, ou seja, $K' = 1$. Seja $\{d_{i1}, d_{i2}, \dots, d_{iK'}\}$ o conjunto das distâncias entre o pixel x_i e as K' médias em M' . Se d_{ij} , a menor destas distâncias, é inferior ao limiar L , a média m_j é atualizada. Se d_{ij} é maior que L , uma nova média, $m_{K'+1} = x_i$, é incorporada a M' . Sempre que a média de alguma classe se modificar, verifica-se se a nova média ainda está a uma distância maior que L da classe mais próxima. Se não estiver, estas classes tornam-se uma única, com média igual à média ponderada das classes que a originaram.

O valor de L , determinado empiricamente, é dado pela metade da soma das variâncias da imagem em cada banda, quando se utiliza como medida de distância o quadrado da distância euclidiana. Este valor pode ser multiplicado por um fator menor que 1, caso o resultado final da primeira etapa apresente um K' menor que K . Para as experiências realizadas, no entanto, os resultados foram sempre satisfatórios, obtendo-se para K' um valor da ordem de 3 vezes o valor de K .

Na segunda etapa, o KMAH opera sobre o conjunto M' de K' médias obtidas na primeira, usando o método de aglomeração hierárquica [Duda - Hart (1973)] para chegar a M .

2.2 Os Classificadores Paramétricos por Pixel

Todos os classificadores paramétricos analisados neste trabalho consideram uma modelagem gaussiana. Portanto, as funções densidade de probabilidade (fdp's) condicionadas às classes podem ser escritas:

$$p(x|h_i) = \frac{1}{(2\pi)^{B/2} |A_i|^{1/2}} \times \exp \left[-\frac{1}{2} (x - m_i)' A_i^{-1} (x - m_i) \right], \quad i = 1, \dots, K \quad (1)$$

onde B é o número de bandas da imagem, m_i e A_i são, respectivamente, o vetor média e a matriz covariância da classe h_i , e K é o número de classes da

imagem. Os parâmetros são estimados na etapa de treinamento.

Pode-se mostrar [Duda-Hart (1973)] que a regra de decisão estatística que leva à menor probabilidade de erro na classificação é:

$$\text{Atribui-se } x \text{ à classe } h_i \text{ se } p(x|h_i)P(h_i) > p(x|h_j)P(h_j) \quad \forall j \neq i$$

Definindo-se as *funções discriminantes*, $g_i(x)$, como

$$g_i(x) = -\ln \{p(x|h_i)P(h_i)\}, \quad i = 1, \dots, K \quad (2)$$

tem-se o seguinte enunciado para o processo de decisão:

$$\text{Atribui-se } x \text{ à classe } h_i \text{ se } g_i(x) < g_j(x) \quad \forall j \neq i$$

No caso específico de modelagem gaussiana, as funções discriminantes podem ser escritas da seguinte forma:

$$g_i(x) = \frac{1}{2} (x - m_i)' A_i^{-1} (x - m_i) + \frac{1}{2} \ln |A_i| - \ln P(h_i) \quad i = 1, \dots, K \quad (3)$$

2.2.1 O Classificador de Máxima Probabilidade a Posteriori

O classificador de máxima probabilidade *a posteriori* (MAP) utiliza exatamente o conjunto de funções discriminantes descrito na Eq. (3). A partir de uma pré-classificação efetuada pelo KMAH, encontra-se as estimativas de máxima verossimilhança para m_i e A_i , e a frequência relativa de cada classe é usada como estimativa para as probabilidades *a priori*

$$P(h_i) = n_i/N \quad (4)$$

onde n_i é o número de pixels atribuídos à classe h_i pelo KMAH, e N é o número de pixels da imagem.

Após a primeira iteração, os valores de $P(h_i)$ utilizados são aqueles encontrados na iteração anterior. Este processo prossegue até que um critério de parada se verifique.

2.2.2 O Classificador MAP Adaptativo

O MAP Adaptativo (MAPA) [Zyl-Burnette (1992)], utiliza as mesmas funções discriminantes que o MAP, porém as probabilidades *a priori* são estimadas pela

frequência relativa de cada classe dentro de uma janela em torno do pixel a ser classificado. O objetivo é explorar mais profundamente as características locais da imagem.

2.2.3 O Classificador de Máxima Verossimilhança

No classificador de máxima verossimilhança (MV), considera-se que as classes são equiprováveis. Assim, as probabilidades *a priori* não interferem na decisão, e as funções discriminantes podem ser simplificadas para:

$$g_i(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{m}_i)' \mathbf{A}_i^{-1} (\mathbf{x} - \mathbf{m}_i) + \frac{1}{2} \ln |\mathbf{A}_i| \quad (5)$$

2.3 O Classificador Contextual

Os classificadores convencionais, pixel a pixel, valem-se unicamente da observação das componentes do pixel \mathbf{x}_{ij} para classificá-lo. Já os classificadores contextuais utilizam a informação contida em uma vizinhança ou contexto, V , de \mathbf{x}_{ij} .

A função discriminante do classificador contextual (CX) proposto por Tilton-Swain [Tilton-Swain (1981)], é definida usando-se um conjunto de p pixels formado pelo próprio \mathbf{x}_{ij} e mais $p - 1$ pixels em V . Este conjunto, denominado *p-contexto* de \mathbf{x}_{ij} , é representado pelo vetor $\mathbf{x}_{ij}^p = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)$, onde \mathbf{x}_p é sempre o pixel a ser classificado e os demais pixels mantêm uma configuração espacial fixa com relação à posição de \mathbf{x}_{ij} . Define-se ainda o *vetor configuração*, p -dimensional, $\mathbf{v}^p = (v_1, v_2, \dots, v_p)$, com $v_k \in \{h_0, h_1, \dots, h_K\}$, ficando entendido que as componentes v_k mantêm entre si a mesma configuração espacial especificada para \mathbf{x}_{ij}^p .

A consideração de independência condicionada à classe para os pixels contidos em V [Tilton-Swain (1981)] permite escrever as funções discriminantes do CX como

$$g_i(\mathbf{x}) = - \sum_{\substack{\mathbf{v}^p \in \Omega^p \\ v_p = h_i}} G(\mathbf{v}^p) \prod_{k=1}^p p(\mathbf{x}_k | v_k) \quad (6)$$

onde Ω^p é o conjunto de todos os vetores \mathbf{v}^p , e $G(\mathbf{v}^p)$, a *função contexto*, é a frequência relativa de ocorrência de \mathbf{v}^p na imagem.

A estimação de $G(\mathbf{v}^p)$ utiliza uma pré-classificação da imagem obtida pelo KMAH, levantando-se o histograma das possíveis configurações de classes espacialmente dispostas na forma de \mathbf{v}^p .

Após a primeira iteração da classificação contextual, estima-se a nova função contexto de acordo

com o resultado obtido, e repete-se o processo até que uma condição de parada se verifique.

3 Descrição dos Classificadores Propostos

3.1 O Classificador MAP por Sub-imagens

O MAP tradicional utiliza estimativas das probabilidades *a priori* feitas sobre a imagem inteira e, de posse dessas estimativas, procura minimizar a probabilidade de erro de classificação. O MAP por sub-imagens (MAPSI), aqui proposto, particiona a imagem original em imagens menores, classificando-as como imagens isoladas. Os parâmetros das gaussianas são aqueles obtidos no treinamento da imagem original, sem particionamento, mas as probabilidades *a priori* são estimadas dentro de cada sub-imagem. Assim, procura-se minimizar a probabilidade de erro de classificação em cada sub-imagem.

3.2 O Contextual por Sub-imagens

O contextual por sub-imagens (CXSI) particiona a imagem a ser classificada em imagens menores, estimando uma função contexto específica para cada uma destas. Da mesma forma que o MAPSI, este classificador foi proposto com o objetivo de explorar as características locais da imagem. O CXSI usa uma pré-classificação realizada pelo KMAH para estimar as diversas funções contexto em cada sub-imagem. Os parâmetros \mathbf{m}_i e \mathbf{A}_i , no entanto, são estimados sobre a imagem original, não particionada.

4 Resultados

Nesta seção são apresentados os resultados de classificação aplicando-se os diversos classificadores descritos a algumas imagens. Duas imagens sintéticas, SIN1 (400 x 400, 3 bandas) e SIN2 (512 x 512, 3 bandas) foram utilizadas para avaliar a taxa de erro de classificação. As imagens sintéticas são imagens com classes conhecidas, geradas a partir de uma imagem de sensoriamento remoto classificada. Adicionando-se ruído gaussiano de forma controlada, produz-se uma imagem que apresenta características semelhantes às de imagens reais.

Na comparação entre os classificadores MAPA e CXSI também foram usadas as duas imagens reais de sensoriamento remoto, AM1 (400 x 400, 3 bandas) e AM2 (512 x 512, 3 bandas) a partir das quais as imagens SIN1 e SIN2 foram geradas. O parâmetro usado nesta comparação não pode ser a taxa de erro de classificação, uma vez que as classes corretas não são conhecidas. Definiu-se desta forma o parâmetro *discrepância*, que mede a porcentagem de pixels classificados diferentemente por dois classificadores.

Nas classificações realizadas neste trabalho dois critérios de parada foram utilizados para o MAP. O primeiro consiste em parar após uma iteração do algoritmo (MAP1) e o segundo após os valores das médias representativas estabilizarem-se (MAP2). O MAPA e o MAPSI usaram apenas o critério de estabilização das médias, alcançados em ambos os casos após 5 iterações. No MAPSI o tamanho das sub-imagens foi fixado em 8 x 8 e no MAPA foram usadas janelas 3 x 3, para a imagem SIN1, e 5 x 5, para a imagem SIN2. O critério de parada para os contextuais foi o de uma iteração c , para o CXSI, foram usadas janelas de dimensões 16 x 16. No p-contexto, para estes classificadores, considerou-se $p = 3$ e a configuração espacial determinada pelas coordenadas (i, j) , $(i - 1, j)$ e $(i, j - 1)$.

Além da taxa de erro (E), as tabelas a seguir apresentam a taxa média de erro por classe (ECMd), a taxa máxima de erro por classe (ECMx), todos em valores percentuais, e o tempo de processamento, em segundos. Todas as imagens foram classificadas em 6 classes.

Os resultados de classificação das imagens sintéticas SIN1 e SIN2 encontram-se nas tabelas 1 e 2, onde se pode ver que os diversos classificadores não apresentam grandes diferenças de desempenho. No caso da imagem SIN2 o MV foi destacadamente inferior, em termos de taxa de erro. A observação da taxa média de erro de classificação por classe apresenta um comportamento semelhante. Pode-se desta forma concluir que o MAPA seria uma boa alternativa, conjugando a menor taxa de erro com um baixo tempo de execução.

Classif.	E	ECMd	ECMx	Tempo
KMAH	2,21	3,50	9,01	108,70
MV	2,50	3,25	8,38	33,89
MAP1	2,70	4,53	9,82	37,45
MAP2	2,70	4,52	9,82	99,75
MAPA	1,64	2,44	6,42	207,82
MAPSI	1,83	2,72	7,15	166,30
CX	1,89	2,55	7,76	1.690,56
CXSI	1,73	2,48	6,94	1.718,19

Tabela 1: Desempenho dos classificadores (imagem SIN1)

Foram realizados testes para verificar o efeito do tamanho da janela no desempenho do MAPA. Uma janela de grandes dimensões se opõe ao objetivo básico do MAPA, aproximando-o de um MAP convencional. Por outro lado, quanto menor for a janela,

Classif.	E	ECMd	ECMx	Tempo
KMAH	5,09	12,94	50,30	229,69
MV	9,93	9,94	30,65	56,14
MAP1	6,81	14,32	44,17	64,27
MAP2	6,88	14,50	42,96	177,29
MAPA	4,38	9,04	35,43	354,41
MAPSI	4,91	10,31	38,07	284,26
CX	4,72	10,99	43,42	2.464,23
CXSI	4,63	10,96	43,92	2.540,69

Tabela 2: Desempenho dos classificadores (imagem SIN2)

menor a precisão na estimativa das probabilidades *a priori*, o que pode se refletir numa degradação do desempenho. Existe, portanto, um compromisso. Zyl [Zyl-Burnette (1992)] sugere uma janela contendo um número de pixels aproximadamente igual a dez vezes o número de classes. Para imagens classificadas em seis classes a janela conteria então aproximadamente 60 pixels. No entanto, os resultados apresentados nas tabelas 3 e 4 mostram que, para as imagens SIN1 e SIN2, o melhor desempenho foi obtido para janelas pequenas, da ordem de 16 pixels.

Janela	E	ECMd	ECMx	Tempo
16 x 16	1,87	2,86	7,26	507,10
8 x 8	1,75	2,62	6,94	258,12
6 x 6	1,69	2,54	6,77	224,95
4 x 4	1,67	2,51	6,62	215,21
3 x 3	1,64	2,44	6,42	207,82
3 x 1	1,82	2,77	6,74	149,04

Tabela 3: Desempenho do MAPA para diferentes janelas (imagem SIN1)

Janela	E	ECMd	ECMx	Tempo
8 x 8	4,59	9,64	36,57	423,94
6 x 6	4,46	9,22	35,70	375,88
5 x 5	4,38	9,04	35,43	354,41
4 x 4	4,37	9,04	36,04	333,85
3 x 3	4,44	8,97	36,13	319,56

Tabela 4: Desempenho do MAPA para diferentes janelas (imagem SIN2)

O desempenho manteve-se praticamente cons-

tante para janelas entre 9 e 64 pixels. Como o custo computacional aumenta com o tamanho da janela, janelas pequenas são uma escolha mais interessante, neste caso.

Para verificar se o desempenho do CXSI é realmente próximo ao do MAPA, calculou-se a discrepância entre os dois na classificação das duas imagens reais, AM1 e AM2. Para a AM1, os classificadores divergiram em 1,38%, enquanto que, para a AM2, a discrepância chegou a 3,96%. Sabendo-se que estes são os valores máximos da diferença de desempenho entre os dois métodos e que o tempo de processamento do CXSI é muito maior que o do MAPA, mais uma vez este surge como a alternativa mais indicada.

5 Referências

- J. J. Van Zyl and C. F. Burnette, Bayesian Classification of Polarimetric SAR Images Using Adaptive a priori Probabilities, *International Journal of Remote Sensing*, vol. 13, no 5. (Mar. 1992) 835-840.
- J. C. Tilton and P. H. Swain, Incorporating Spatial Context into Statistical Classification of Multidimensional Image Data, *Laboratory for Applications of Remote Sensing*, Purdue University (Aug. 1981).
- R. O. Duda and P. E. Hart, *Pattern Classification And Scene Analysis*, John Wiley & Sons, 1973.