

Classificação hierárquica orientada a objeto em imagens de alta resolução espacial empregando atributos espaciais e espectrais

Marlos Henrique Batista
Vitor Haertel

Universidade Federal do Rio Grande do Sul - UFRGS/CEPSRM
Caixa Postal 15.044 – 91501-970 – Porto Alegre - RS, Brasil
marlos.henrique@ufrgs.br; vitor.haertel@ufrgs.br

Abstract. The traditional pixel based classification procedures frequently yield poor results when applied to high spatial resolution image data. The small ground pixel size frequently captures an unnecessary high level of detail resulting in large class-variances which increase the confusion among classes. In this context, it is normally more convenient to replace individual pixels by image segments, also known as objects, as individuals to be labeled according with the existing classes. One of the main advantages of employing smaller pixels is the ability of capturing spatial information such as shape and texture which may be conveniently added to the spectral data to increase the accuracy in the resulting thematic map. In this study, it is investigated the combined use of spectral and spatial attributes in a classifier. As the data dimensionality may increase substantially due to the larger amount of data used, a feature extraction process are used in the classification procedure, rather than the original spectral and spatial data themselves. A binary decision tree classifier is used in this study to label every image segment.

Palavras-chave high spatial resolution, hierarchical classifier, object-oriented classification, imagem de satélite de alta resolução espacial, classificação orientada a objetos em estágios múltiplos.

1. Introdução

Pela natureza e composição dos dados imagem em alta resolução espacial, ascende-se a possibilidade de introdução no processo de classificação de atributos de ordem espacial. O principal problema dessa utilização verifica-se na captura excessiva de variações espectrais internas das classes, funcionando como ruído e aumentando desnecessariamente a variância interna das classes. Estas características geram prejuízos no processo de classificação, pois aumentam o erro de *Bayes* e conseqüentemente pioram a acurácia da classificação. Por isso, a necessidade de abandonar o *pixel* como unidade de classificação e passar a adotar regiões (objetos) obtidas por segmentação prévia da imagem. Convergindo a composição dos dados imagem e a necessidade de adoção de objetos, num processo de classificação, investigou-se a utilização de atributos geométricos e de textura em adição aos atributos espectrais.

Para poder avaliar a eficácia de cada atributo ou de um conjunto de atributos no processo de classificação, utilizou-se um algoritmo de classificação em múltiplos estágios, em especial o classificador de decisão em árvore (CDA). A estrutura do CDA apresentada está na forma de uma hierarquia binária, de modo que somente duas classes sejam consideradas em cada estágio (nó). Essa estrutura apresenta a vantagem de permitir uma seleção mais adequada para dados espectrais e espaciais em cada um dos nós. Em lugar de selecionar o conjunto de “bandas” espectrais / espaciais que ofereçam a maior separabilidade entre a totalidade das classes, como em um classificador em estágio único, um CDA binário permite que em cada nó seja utilizado um conjunto próprio de bandas, aquele que apresentar a maior separabilidade entre as duas classes tratadas no nó. Nesta abordagem, para a escolha do par de classes a ser tratado em cada nó individual do CDA, utilizou-se o critério de máxima separabilidade, isto é, entre os vários possíveis pares de classes optou-se por aquele em que as classes apresentavam a maior separação, conforme estimada pela distância de *Bhattacharyya*.

Para utilização conjunta e otimizada dos atributos espectrais e espaciais no processo de classificação, foram selecionados por um processo de seleção de feições denominado SFS

(*Sequential Forward Selection*) um subconjunto de feições em cada nó, que melhor caracterizam o par de classes que apresentar a maior separabilidade encontrada pelo critério da distância de *Bhattacharyya* máxima.

A aplicação do algoritmo proposto foi testado sobre os dados imagem do sistema *Quickbird*, com as bandas multiespectrais e pancromática.

2. Classificadores de Decisão em Árvore

Classificadores de decisão em árvore (CDA) podem ser vistas como um modo de representar um conjunto de regras sobre uma seqüência hierárquica, estruturada de modo a particionar recursivamente os dados. Os CDA's são formados por n -ésimas ramificações de uma árvore, que representam regras de classificação, definidas para classificar os padrões de um determinado domínio em uma configuração de classes mutuamente exclusivas.

A proposta básica em qualquer abordagem em estágios múltiplos é a desagregação de regras de decisões complexas em um conjunto de decisões simples, de modo que, o resultado se mantenha inalterado, sendo qualquer que seja o modelo de decisão adotado.

A estrutura geral de um CDA segue conforme a definição de Safavian e Landgrebe (1991) na **Figura 1**.

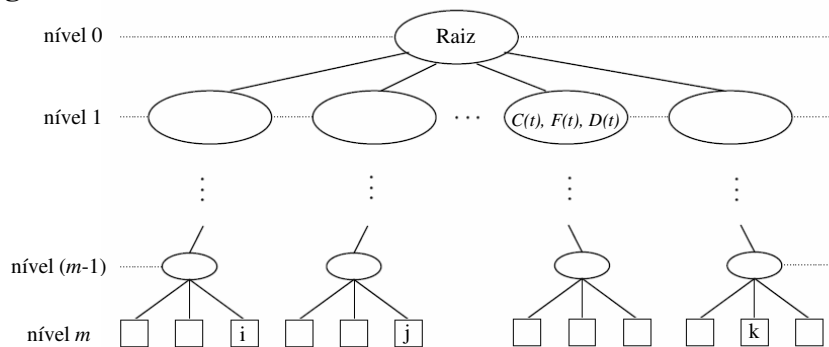


Figura 1 – Estrutura geral de um classificador de decisão em árvore (Adaptado de Safavian e Landgrebe (1991))

No nível 0 está localizado o nó raiz, o início da árvore. O nó raiz contém todos os padrões de todas as n classes a serem classificados pelo CDA. Cada nó seguinte está aqui representado por t , que representa uma terna $(C(t), F(t), D(t))$, onde $C(t)$ representa as classes do nó t , $F(t)$ as feições usadas pelo nó t e $D(t)$ a regra de decisão do nó t . O processo de discriminação dos padrões no CDA utiliza em cada nó uma regra de decisão sobre um conjunto de feições para tentar discriminar um determinado conjunto de classes. Esse processo pode ser considerado como o crescimento da árvore. Os nós apresentados pelo nível m , onde m é o número total de níveis, são chamados de nós terminais. No nó terminal o padrão agora discriminado, recebe a identificação da classe do nó.

Uma vantagem expressiva que o CDA apresenta é a permissão de que classes apresentadas em estágios intermediários do processo de classificação sejam rejeitadas, fazendo com que um padrão seja exposto a um número reduzido de classes, diminuindo o erro de *Bayes*, otimizando o desempenho do classificador e o tempo de processamento. Outras potencialidades foram apresentadas por Safavian e Landgrebe (1991) em referência ao CDA.

3. Segmentação Multiresolução

A segmentação multiresolução pode ser descrita como uma técnica *bottom-up* de fusão de regiões, que leva em consideração múltiplas características da imagem e que utiliza um algoritmo baseado em crescimento de regiões (Baatz e Schape, 2000). O processo de segmentação inicia-se considerando cada *pixel* como um objeto. Iterativamente um par de

objetos é agrupado em um objeto de dimensão maior. A regra para o agrupamento (fusão) é baseada em um critério de heterogeneidade, que descreve a dissimilaridade de objetos adjacentes. O critério de heterogeneidade é uma combinação de propriedades espectrais e espaciais. O algoritmo descreve estas propriedades como critérios de cor (espectrais) e de forma (espaciais). O critério de forma ainda é subdividido entre os parâmetros de compacidade e suavidade. Estes critérios podem ser traduzidos como uma ponderação em que a segmentação buscará priorizar na formação dos objetos.

Considerando um objeto arbitrário A , adjacente a outros n objetos, o par de objetos que apresentar o menor crescimento no critério de heterogeneidade é agrupado, em outras palavras, será agrupado o par de objetos que apresentar a maior similaridade. O processo termina quando o menor crescimento possível de um par de objetos excederem certo limiar, o parâmetro de escala. Um maior parâmetro de escala permitirá um número maior de iterações e agrupamentos, conseqüentemente resultará em objetos com maior dimensão.

Os principais componentes da segmentação multiresolução, conforme Baatz e Schape (2000) são: decisão heurística para determinar os objetos que serão agrupados em cada etapa; definição do critério de heterogeneidade de um par de objetos.

A decisão heurística no processo de segmentação multiresolução cumpre o papel de encontrar os objetos que serão agrupados em cada iteração. O critério de heterogeneidade busca definir um indicador da similaridade entre dois objetos considerando todas as n -dimensões do domínio espacial.

4. Atributos para Caracterização dos Objetos

Com a proposta de incrementar variáveis (atributos) no processo de classificação, sobre a utilização combinada, modelou-se atributos de natureza espectral e espaciais sobre os objetos definidos no processo de classificação hierárquica, promovendo a determinação de parâmetros que quantifiquem as eficiências de modo individual e conjuta. Os atributos espectrais propostos caracterizam a radiância refletida pelos objetos de interesse nas diversas regiões do espectro eletromagnético abrangidas pelos dados. Este atributo define o comportamento espectral de cada objeto em particular. As variáveis espectrais foram definidas segundo as faixas espectrais de 0.45 a 0.52 μm , 0.52 a 0.60 μm , 0.60 a 0.69 μm , 0.76 a 0.90 μm , e são: média espectral; desvio-padrão espectral; razão espectral; brilho; e diferença máxima.

Os atributos espaciais foram classificados em termos geométricos (ou de forma) e de textura. Os atributos espaciais geométricos são aqueles empregados pela morfologia matemáticas dos elementos e modelou-se os atributos: área; comprimento; largura; comprimento por largura; compactação; ajuste elíptico; ajuste retangular; comprimento da borda; índice de forma; densidade; direção principal; e assimetria.

Os atributos espaciais de textura foram definidos pela utilização de estatísticas de segunda ordem, estimadas a partir das matrizes de coocorrência dos níveis de cinza (*Grey Level Cooccurrence Matrix*). As variáveis de textura aqui modeladas seguiram a proposta de Haralick (1973), compreendendo as GLCM de: homogeneidade, contraste, dissimilaridade, entropia, média, desvio-padrão e correlação.

5. Utilização Conjunta de Atributos Espectrais e Espaciais no Processo de Classificação

Dado um conjunto de variáveis, freqüentemente ocorre que algumas variáveis são altamente correlacionadas, não contribuindo para incrementar o poder de discriminação entre as várias classes, mas aumentando desnecessariamente a dimensionalidade dos dados e com isto, o número de parâmetros a serem estimados. Esta situação torna-se especialmente nociva, quando o número de amostras de treinamento é pequeno, levando a uma degradação na performance do classificador (fenômeno de *Hughes*).

De um modo geral, existem dois processos para atingir esta finalidade: seleção de feições e extração de feições. Como aqui investigado, o método de seleção de feições procura selecionar o subconjunto mais adequado de variáveis, no sentido de maximizar a separação entre as classes conforme um critério previamente estabelecido. O critério da distância de *Bhattacharyya* foi adotado para selecionar o par de classes mais distantes em cada nó da árvore. A distância de *Bhattacharyya* fornece uma medida de separabilidade entre duas classes. Ela é caracterizada como ótima quando utilizam-se um par de classes normais e sub-ótima para situações envolvendo mais de duas classes ao mesmo tempo (Fukunaga, 1990). Na suposição de normalidade dos dados a distância de *Bhattacharyya* fica:

$$b_{ij} = \frac{1}{8} \cdot (M_i - M_j)^T \cdot \left(\frac{\Sigma_i + \Sigma_j}{2} \right)^{-1} \cdot (M_i - M_j) + \frac{1}{2} \cdot \ln \left\{ \frac{|\Sigma_i + \Sigma_j|/2}{\sqrt{|\Sigma_i| |\Sigma_j|}} \right\}$$

Para identificar iterativamente o melhor subconjunto de feições que possa ser obtido pela adição de uma feição por iteração, ao invés de selecionar um conjunto completo de feições numa única iteração, investigou-se a Seleção Seqüencial Progressiva (*Sequential Forward Selection* - SFS), que é um método *bottom-up* de seleção de feições. Uma particularidade que o SFS apresenta é que a feição uma vez selecionada não pode ser descartada, o que o torna sub-ótimo.

6. Experimentos

A definição das classes de informação e a investigação do comportamento e das aproximações quanto às distribuições das classes, segundo o espaço multivariado, estruturado por objetos, propuseram uma preliminar seleção das variáveis que seriam submetidas aos experimentos realizados sobre o CDA. Os experimentos iniciais do CDA foram delineados a partir da variação do número de feições utilizadas em cada nó da árvore, partindo da utilização de todo conjunto de amostras de treinamento disposto. Nesta etapa, investigou-se as características das árvores geradas, conforme a variação de 5, 10, 15 e 20 feições por nó, sobre a utilização conjunta das variáveis espectrais e espaciais. Sendo 31 variáveis propostas, 14 variáveis espectrais e 17 variáveis espaciais. Na metodologia do CDA somente as classes que cumprirem certo critério de pertinência serão herdadas para os nós descendentes. Esse critério apresenta o conceito de limiar de verossimilhança, que representa um fator mínimo de verossimilhança que certa classe deve apresentar para ser classificada no nó descendente. Os demais experimentos foram partidos da investigação das árvores segundo a variação dos limiares de verossimilhança (LV) de 80%, 90%, 95% e 99%. Sobre cada LV foram utilizadas as mesmas variações no número de feições por nó, resultando numa totalidade de 16 árvores. Em todos os experimentos foram utilizadas sete classes. A avaliação da acurácia individual de cada classe, para cada árvore, foi auferida pelo método *hold-out*, através das amostras de testes obtidas por simples seleção nos dados imagem. Essa avaliação permitiu a determinação do melhor CDA, ou seja, daquela que produza a melhor acurácia para um conjunto de classes. Utilizou-se como medidas de acurácia, o índice *kappa*, a acurácia média do classificador e do usuário.

Os experimentos de segmentação e a geração das variáveis dos atributos foram realizadas através do software comercial *eCogniton*. A implementação dos demais experimentos foram realizadas através do desenvolvimento de um algoritmo em ambiente *Matlab*.

6.1. Experimentos de Segmentação

Os experimentos de segmentação foram constituídos pelo método multiresolução, através do software comercial *eCognition*. Esses experimentos partiram da utilização da segmentação em

níveis, através de combinações das variáveis de entrada, constituindo ao total 18 níveis de dados segmentados.

A proposta dos experimentos de segmentação é o de encontrar o nível de segmentação que melhor representem os limites dos elementos reais apresentados nos dados imagem original, estruturados com o menor efeito possível de representação de *over-segmentation* e *under-segmentation*. A **Figura 2** apresenta um pequeno recorte dos dados imagem segmentados em quatro níveis, nível 5, 6, 7 e 8.

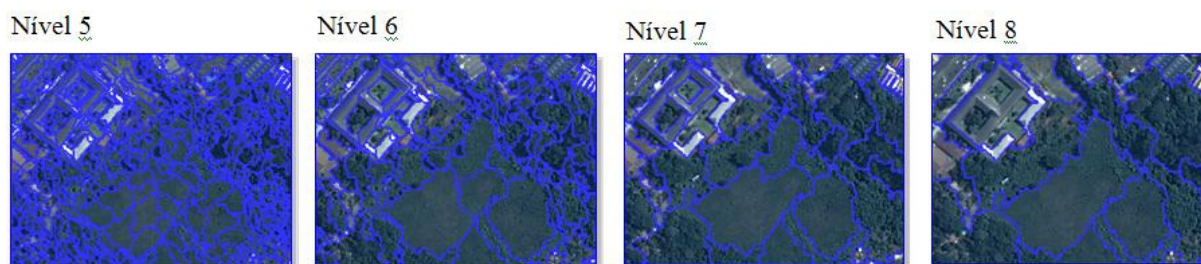


Figura 2 – Níveis de segmentação do primeiro conjunto de experimentos

Alguns critérios de seleção foram adotados para obtenção dos dados imagem segmentado, tais como, análise visual, estatísticas da imagens, comparações com outras fontes de dados (ex. restituição fotogramétrica) e relação à borda dos elementos reais apresentados pelos dados imagem original. Dos níveis gerados, selecionou-se o Nível 6 como os dados imagem segmentados mais adequado.

6.2. Seleção Preliminar das Variáveis e Classes de Informação

O principal critério para seleção das variáveis e das classes de informação, que serão dispostas ao CDA são a normalidade dos dados e a quantidade de amostras, respectivamente. A normalidade dos dados foi verificada por aproximação visual através de análise dos histogramas. A quantidade de amostras de treinamento e de testes para cada classe foi o critério seleção das classes de informação, sendo que apenas a classe hidrografia forneceu uma quantidade praticamente inexpressiva para obtenção das estatísticas da classe e por consequência à definição de um padrão.

As assinaturas (espectrais/espaciais) das classes selecionadas, utilizando o valor média e o desvio-padrão de cada variável, foram dispostas em gráficos para verificar o comportamento geral (individual e combinado) dos atributos e das classes. A imagem apresentada na **Figura 3** apresentam estas assinaturas, onde no eixo horizontal estão representados pelos atributos e no eixo vertical estão representados os valores de cada atributo, padronizado em 8 bits.

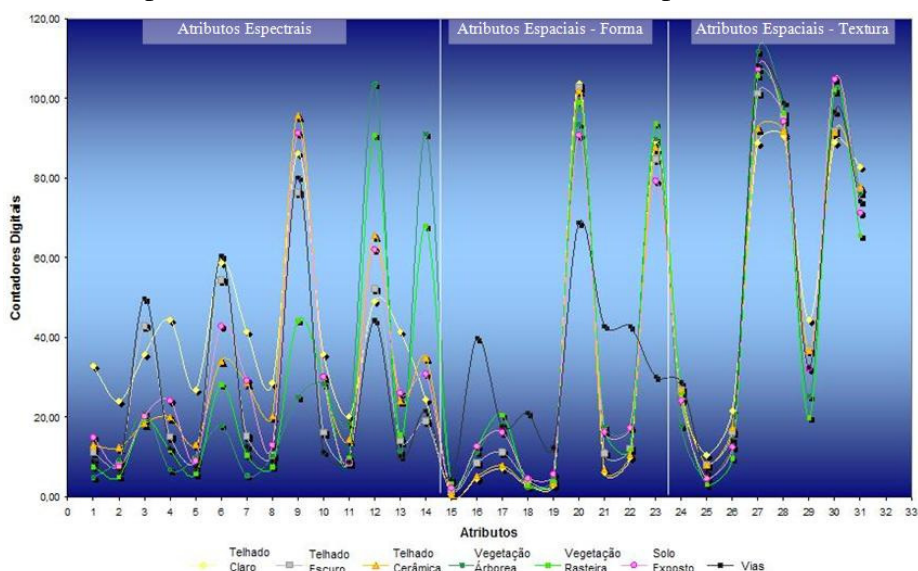


Figura 3 – Assinaturas espectral/espacial para as sete classes de informação

6.3. Experimentos do Classificador de Decisão em Árvore

Foram realizados 16 conjuntos de experimentos com os CDA, aplicando um incremento quanto ao número de variáveis por nó, alterações do limiar de verossimilhança, avaliando o comportamento da árvore, das classes e variáveis selecionadas e a acurácia da classificação. Dos experimentos do CDA o que apresentou a melhor acurácia geral, continha um LV fixo de 99% e possuía 20 feições em cada nó. A estrutura deste CDA compreende 51 nós, sendo deste, 26 nós terminais. A **Figura 4-A** apresenta o conjunto de classes de cada nó, sendo o par de classes com maior separação apresentado pela cor em destaque e a **Figura 4-B** apresenta o conjunto de atributos selecionados pelo SFS em cada um destes nós.

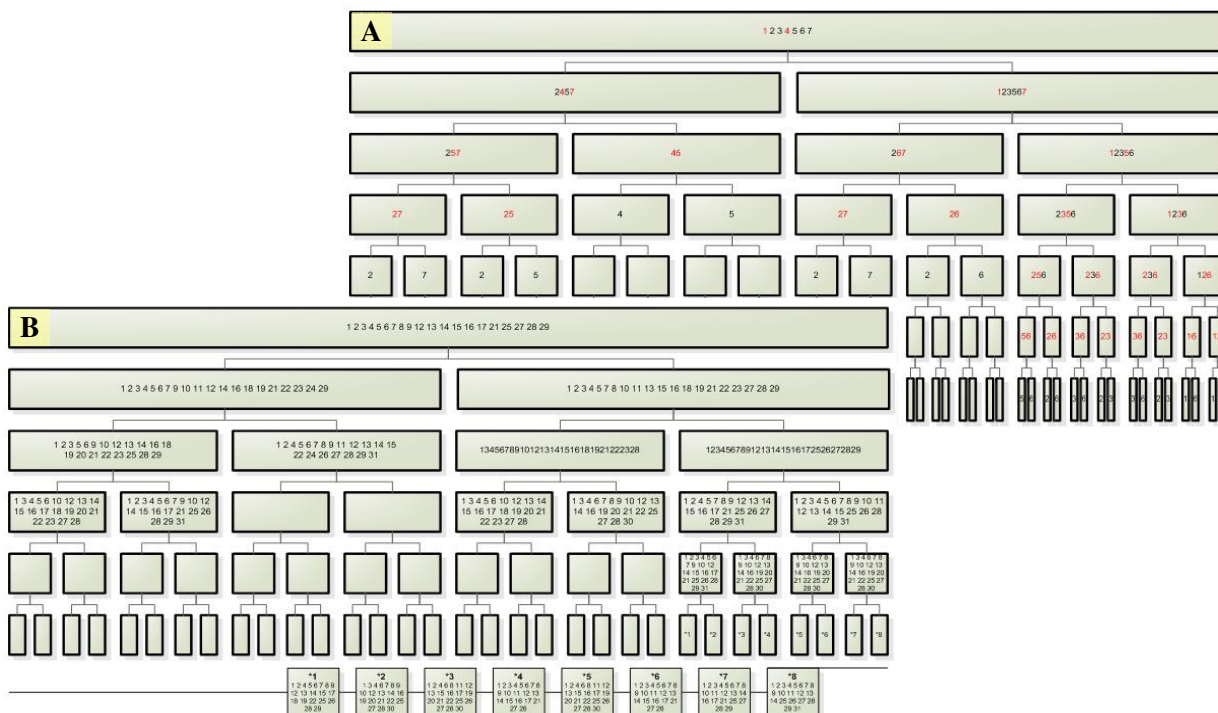


Figura 4 – Estrutura de classes e variáveis do CDA com LV de 99% e 20 feições em cada nó

Das variáveis selecionadas pelo experimento em média 63% são de natureza espectral e 37% de natureza espacial. De todos os conjuntos de experimentos este é o CDA que mais utilizou variáveis espaciais. Para estimação da acurácia do CDA classificaram-se as amostras de testes das classes (método de estimação: *hold-out*). Os erros envolvidos nos experimentos dizem respeito principalmente à confusão entre as classes telhado de cerâmica (3) e solo exposto (6), onde a maioria dos objetos que deveriam ser classificados na classe 3 foram classificados na classe 6.

A acurácia geral estimada neste experimento (índice *kappa*, acurácia média do produtor e acurácia média do usuário) oscilou em torno de 88%, apresentando uma acurácia muito similar para todas as árvores deste conjunto de experimentos. A acurácia do usuário manteve-se ligeiramente superior à acurácia do produtor, como nos demais experimentos, indicando a ocorrência de um número maior de erros de inclusão do que de omissão. Dos conjuntos de experimentos algumas considerações podem ser relevantes: o experimento com maior número de nós está com 59 nós; a uma tendência de seleção dos atributos espaciais quando há a presença, principalmente das classes 2, 3, 6 e 7; o principal fator de erro neste conjunto de experimentos é devido à inclusão de objetos que deveriam ser classificados na classe 3 e foram classificados na classe 6; a acurácia do usuário para todas as classes, com exceção da classe 3, foi superior à 95%; a acurácia do classificador para todas as classes, com exceção da classe 6, foi superior à 90%.

6.4. Classificador de Decisão em Árvore x Classificador em Estágio Único

Um conjunto adicional de experimentos teve por objetivo fazer uma comparação entre o classificador em estágio múltiplo adotado (CDA) e um classificador tradicional em estágio único (CEU), este último utilizando *pixels* individuais como padrões a serem classificados. Nestes experimentos, o classificador em estágio único adotado foi o da Máxima Verossimilhança Gaussiana (MVG).

Para proposta de comparação visual da classificação neste experimento, resolveu-se utilizar todos os dados imagem disponíveis e adotou-se uma única árvore para CDA, a que proporcionou a melhor acurácia dentre os 16 CDA investigados, com LV de 99% e 20 feições em cada nó. A **Figura 5-A** corresponde aos dados imagem original do sistema *Quickbird* apresentado em cor verdadeira pela composição RGB (3,2,1). A **Figura 5-B** ao resultado do CDA e a **Figura 5-C** ao resultado do CEU.

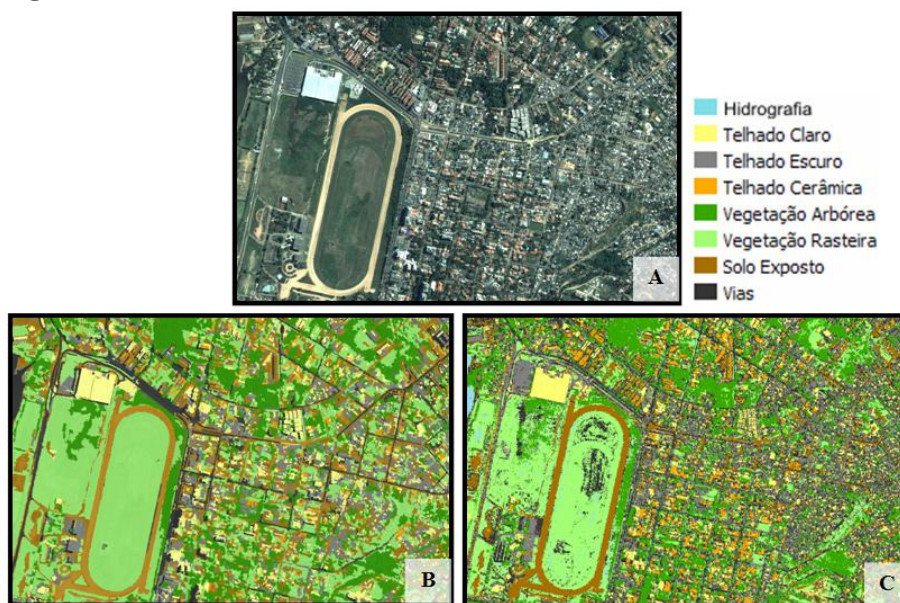


Figura 5 – Resultado gráfico do CEU e do CDA

Por análise da tabela de contingência, pode-se inferir que os erros envolvidos no CEU não podem ser pontuados conforme os apresentados pelo CDA. Neste caso não há uma tendência de comportamento (erros de omissão e/ou inclusão) sobre determinadas classes. O que realmente pode-se verificar é a presença de erros de omissão e de inclusão apresentados de modo distribuído. Os principais erros apresentados pelo CEU são de omissão e inclusão quanto às classes telhado (2) escuro e vias (7).

7. Resultados

Verificou-se que, pelas acurácias gerais, o CDA com o maior número de variáveis (20) e com o maior limiar de verossimilhança (99%) encontrou a menor incidência de erros entre todos os experimentos propostos, inclusive em comparação com o classificador em estágio único.

Como o objeto principal da pesquisa visa avaliar a eficiência do CDA com diferentes composições (variáveis e LV), estruturou-se gráficos indicando as acurácias de cada árvore e do CEU conforme a acurácia média do classificador. O eixo vertical dos gráficos apresenta a acurácia percentual e o eixo horizontal o número de variáveis utilizadas. Como são quatro variáveis (atributos espectrais) utilizadas pelo CEU, representou-se este, nesta posição. Cada conjunto de experimentos está representado em uma linha do gráfico. A **Figura 6** apresenta o gráfico das medidas de acurácia média do classificado para o CDA e o CEU.

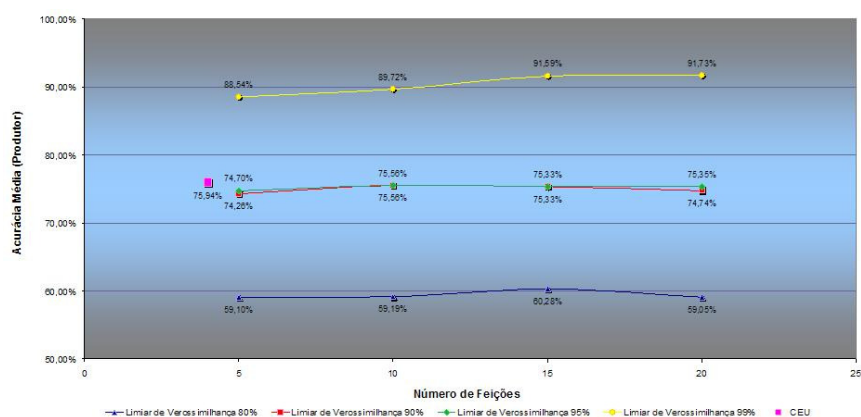


Figura 6 – Gráfico de avaliação da acurácia média do classificador entre os experimentos do CDA e o CEU

8. Conclusão

Os experimentos mostraram um ganho significativo na acurácia das imagens temáticas produzidas pelo método proposto (CDA), com relação ao método tradicional (CEU). A acurácia da classificação pela metodologia proposta atingiu índices aceitáveis e significativamente superiores ao CEU, principalmente quanto maior é a utilização conjunta dos atributos espectrais e espaciais, e nos experimentos com maior LV.

A seleção dos atributos espaciais se dá constantemente em menor número, porém quando há maior presença dos atributos espaciais, há também uma maior acurácia da classificação, indicando a alta correlação entre as variáveis espaciais e ao mesmo tempo um alto poder discriminante dessas variáveis.

Com o aumento do número de feições, para os experimentos de um mesmo LV, não há aumento significativo da acurácia geral (índice *kappa*, acurácia média do produtor e acurácia média do usuário), há um aumento gradual na acurácia, seguindo os efeitos do fenômeno de *Hughes*. Entretanto, ao aumentar o LV, há um aumento expressivo na acurácia geral da classificação. Um menor valor no LV indica um processo de verossimilhança mais flexível, de modo que a classe em teste sofra mais facilmente o efeito de *pruning* (poda) para os demais nós na árvore. Um maior valor do LV irá possibilitar que aos nós descendentes somente as classes que forem mais verossímeis sejam unicamente selecionadas, caso contrário, a classe em teste segue para os dois nós descendentes.

O CDA no processo orientado a objetos empregando atributos espectrais e espaciais permite na grande maioria das vezes uma maior acurácia na classificação, do que aquela que seria possível com a utilização sobre os processos tradicionalmente empregados em dados imagem de sensoriamento remoto, como é o caso dos CEUs baseado a *pixels*. Adicionalmente, os CDA permitem a identificação direta dos erros envolvidos sobre as classes, apresentando frequentemente uma fonte de tendência, diferente ao CEU, que os erros envolvidos apresentam-se distribuídos de modo não ordenado. A tendência neste caso é altamente positiva, pois pode proporcionar o refinamento e o ajuste dos erros no processo de classificação, além de possibilitar a utilização individual das classes que estão inertes aos erros envolvidos.

Ao avaliar a capacidade de incorporação dos resultados da classificação num ambiente SIG (Sistemas de Informações Geográficas) através do software *ArcGIS*, encontrou-se uma solução interoperacional eficaz, não apresentando qualquer deficiência em ambos os processos de comunicação (importação e exportação), permitindo realizar todos os relacionamentos necessários, mantendo constantemente o formato original das informações, de modo que a integração Sensoriamento Remoto x SIG seja compreendida.

Referências Bibliográficas

- Baatz, M.; Schape, A. Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. In: AGIT-Symposium Salzburg. Karlsruhe, 1999.
- Blaschke, T.; Lang, S.; Lorup, E.; Strobl, J.; Zeil, P. Objected oriented image processing in an integrated GIS / remote sensing environment and perspectives for environmental application. **In environmental Information for Planning**, vol. 2, Metropolis-Verlag, 2000.
- Bow, S. T. **Pattern Recognition and Image Preprocessing**. Marcel Dekker, 1992.
- Breiman, L.; Friedman, J. H.; Olshen, R. A.; Stone, R. A. **Classification and Regression Trees (CART)**. Wadsworth Int., Belmont, 1984.
- Cressie, N. A. C. Statistics for spatial data revised edition. Wiley - **Interscience publication**, cap. 7.4, p. 499-534, 1993.
- eCognition. **User's Guide**. Definiens Imaging. Muchen, Germany, 2003.
- Fukunaga, K. **Introduction to Statistical Pattern Recognition**. Ed. Boston: Academic Press, 2nd edition, 1990.
- Haralick, R. M.; Shanmugam, K.; Dinstein, I. Textural features for image classification. **IEEE Trans. on Systems, Man, and Cybernetics**, vol. SMC-3, n. 6, p. 610-621, 1973.
- Haralick, R. M. Statistical and structural approach to texture. **Proceedings of the IEEE**, vol. 67, n. 5, p. 786-803, 1979.
- Hughes, G. F. On the mean accuracy of statistical pattern recognizers. **IEEE Trans. on Information Theory**, vol. IT-14, p. 55- 63, 1968.
- Kurzynki, M.W. The optimal strategy of a tree classifier. **Pattern Recognition Letter**, vol. 1, p. 305-310, 1983.
- Mueller, M.; Segl, K.; Kaufmann, H. Edge- and region-based segmentation technique for the extract of large, man-made objects in high-resolution satellite imagery. **Pattern Recognition**, vol. 37, p.1619-1628, 2004.
- Myint, S. W.; Lam, N. A study of lacunarity-based texture analysis approaches to improve urban image classification. **Computers, Environment and Urban Systems**, vol. 29, p. 501-523, 2005.
- Pavuluri, M. K.; Ramanathan, S.; Daniel, Z. A rule-based classifier using classification and regression tree (CART) approach for urban landscape dynamics. In: International Geoscience and Remote Sensing Symposium, vol. 2, p. 24- 28, 2002.
- Quinlan, J. R. Decision Trees and Decisionmaking. **IEEE Trans. Systems, Man and Cybernetics**, vol. 20, n. 2, p. 339-346, 1990.
- Rounds, E. A combined non-parametric approach to feature selection and binary decision tree design. **Pattern Recognition**, vol. 12, p. 313-317, 1980.
- Ryherd, S.; Woodcock, C. Combining spectral and texture data in the segmentation of remotely sensed images. **Photogrammetric Engineering & Remote Sensing**, vol. 62, n.2, p. 181- 194, 1996.
- Safavian, S. R.; Landgrebe, D. A Survey of Decision Tree Methodology. **IEEE Trans. Systems, Man and Cybernetics**, vol. 21, n. 3, 1991.
- Schowengerdt, R. A. **Remote Sensing, Models and Methods for Image Processing**. Academic Press, 2nd edition, Tucson, 1997.
- Sethi, I. K.; Sarvarayudu, G. Hierarchical classifier design using mutual information. **IEEE Trans. on Pattern Analysis and Machine Intelligence**, vol. 4, n. 4, p. 441-445, 1982.
- Simard, M.; Saatchi, S. S.;Grandi, G. D. The use of decision tree and multiscale texture for classification of JERS-1 SAR data over tropical forest. **IEEE Trans. on Geoscience and Remote Sensing**, vol. 38, n. 5, p. 2310-2321, 2000.
- Wu, C.; Landgrebe, D. A.; Swain, P. The decision tree approach to classification. Purdue University, W. Falayette, REP. RE-EE 75-17, 1975.
- Zhang, D.; Lu, G. Review of shape representation and description techniques. **Pattern Recognition**, vol. 37, p. 1-19, 2004.