

Uso de mineração de dados na identificação da cobertura vegetal através de modelos espectrais: uma contribuição metodológica para o mapeamento na escala 1:100.000

Maíra Vieira Zani¹
Luana Santos do Rosário¹
Rafael Silva de Barros¹
Carla Bernadete Madureira Cruz¹

¹Universidade Federal do Rio de Janeiro – UFRJ
Instituto de Geociências
Av. Athos da Silveira Ramos, 149. Prédio do Centro de Geociências Matemáticas e da
Natureza, Bloco F, Ilha do Fundão.
Cidade Universitária
21941-909 – Rio de Janeiro, RJ – Brasil
{mairazani.geo; lulurosario; barros.rafael; carlamad} @gmail.com

Abstract. This work aims to establish a spectral model that facilitates the identification and mapping of forest areas through classification of satellite images acquired on the West Zone of the city of Rio de Janeiro, a mesoscale approach (1:100,000). The study area in question presents a framework of intense urban growth, both by large commercial enterprises as property which causes the remaining forest areas suffer increasing pressure. Therefore it is extremely important to making maps of vegetation, so that monitoring is done, making it easier to identify and protect the area. To perform this study we used images of TM sensor of Landsat-5 for the year 2010 and the methodology follows the approach of object-based image analysis with the aid of data mining (Weka 3.2) and eCognition software. To validate the results were generated random points in the image that then were analyzed using the Google Earth program, which allowed the generation of a confusion matrix to evaluate the quality of the generated mapping. Importantly, the ranking generated despite not having undergone any manual editing or changing model achieved a high level of quality.

Palavras-chave: remote sensing, forest, object-based image analysis, data mining, sensoriamento remoto, floresta, classificação baseada em objeto, mineração de dados.

1. Introdução

As diversas mudanças no uso da terra em decorrência do desmatamento trazem como consequência graves impactos com relação à manutenção da biodiversidade, visto que a redução das áreas de floresta influenciam negativamente o desenvolvimento dos organismos que dependem desse ambiente para sobreviver. Nesse contexto, é possível destacar a Mata Atlântica como um dos biomas mais ameaçados devido ao seu alto grau de antropização. Segundo um estudo realizado pelo Projeto de Conservação e Utilização Sustentável da Diversidade Biológica Brasileira / Ministério do Meio Ambiente (PROBIO/MMA - 2006) existe menos de 20% de remanescentes florestais espalhados ao longo do território nacional, em diferentes estados de conservação.

Dessa forma é de extrema importância a geração contínua de mapeamentos da cobertura vegetal, de modo a propiciar um efetivo monitoramento, tornando mais fácil a identificação, a proteção e/ou a recuperação dos remanescentes desse bioma. É nesse contexto de monitoramento que o sensoriamento remoto se apresenta como uma importante ferramenta, já que possibilita a aquisição de informações sobre diversos alvos na superfície terrestre em diferentes intervalos de tempo, permitindo assim uma análise multitemporal da vegetação.

Uma das metodologias que vem sendo muito utilizada nos mapeamentos de uso e cobertura da terra é a classificação baseada em objetos, que permite a utilização de um vasto número de parâmetros na distinção das classes a serem mapeadas. Entre esses parâmetros, podemos citar: informações espectrais, contextuais, texturais e quanto à forma dos objetos.

Logo, a escolha dos melhores parâmetros que descreve cada classe se torna muito difícil e tende a piorar de acordo com o grau de complexidade do mapeamento.

Assim, a realização desse trabalho se justifica pela necessidade de automatização do processo de mapeamento da cobertura vegetal em mesoescala, facilitando e agilizando essa etapa. Além da elaboração de novas metodologias na tentativa de minimizar as dificuldades apresentadas pelas metodologias convencionais de mapeamento de áreas de floresta.

A área de estudo, figura 1, abrange a Zona Oeste da cidade do Rio de Janeiro, formada pelas áreas de planejamento 4 e 5, compreende 39 bairros, corresponde a mais da metade da área do município e conta com aproximadamente três milhões de habitantes (IBGE – CENSO 2010). Sobre as áreas de planejamentos mencionadas anteriormente é importante ressaltar que a área de planejamento 4 (AP4) destaca-se por um crescimento muito acima da média para o município do Rio de Janeiro, ultrapassando também a área de planejamento 5 (AP5). A Zona Oeste, principalmente durante os últimos anos, apresenta um quadro de grande crescimento, tanto por parte de grandes empreendimentos comerciais quanto imobiliários, o que faz com que as áreas de remanescentes florestais sofram cada vez mais pressão.

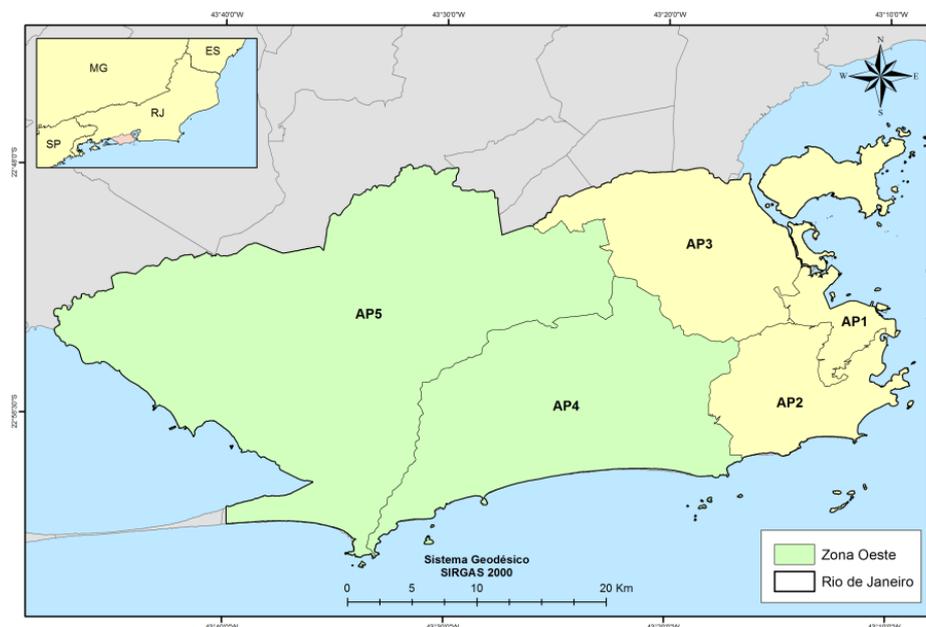


Figura 1. Localização da Área de estudo.

1.2. Objetivo

Analisar a contribuição do uso de descritores espectrais, extraídos por mineração de dados, na identificação e mapeamento de tipos de cobertura vegetal, figura 2, através de análises baseadas em objetos através de imagens LANDSAT 5, adquiridas sobre a Zona Oeste do município do Rio de Janeiro, numa abordagem de mesoescala (1:100.000).

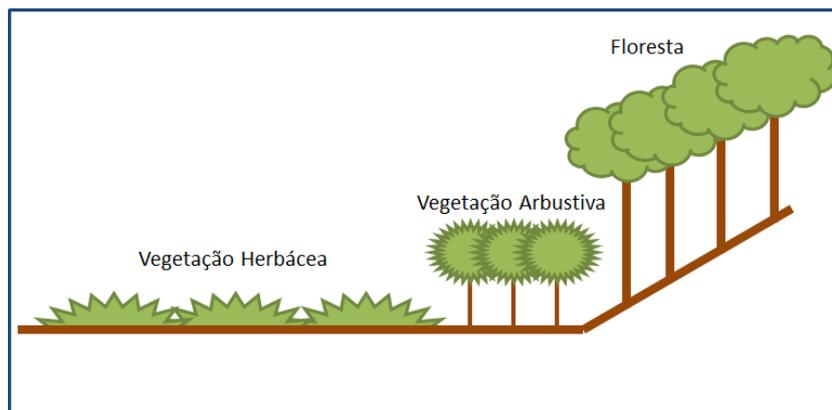


Figura 2. Três níveis de vegetação contemplados pelo mapeamento.

2. Metodologia de Trabalho

Para elaboração do trabalho foi adquirida uma imagem Landsat-5 TM, referente ao ano de 2010. Em seguida foi realizada a correção geométrica da imagem, usando como referência dados vetoriais de ruas da cidade do Rio de Janeiro disponibilizado pelo Instituto Pereira Passos (IPP). O mapa de cobertura e uso da terra foi gerado segundo a abordagem da classificação baseada em objetos do programa Definiens/Recognition, onde foram executadas a segmentação, coleta de amostras para cada classe e escolha de uma lista preliminar de descritores. Para a amostragem das classes foi necessário desmembrar uma mesma classe em várias subclasses de acordo com diferenças observadas na resposta espectral dos alvos, como observado na tabela 1.

Tabela 1. Classes de amostra e respectivas classes de origem.

Classe	Legenda
Floresta_Escura	Floresta
Floresta	Floresta
Floresta_Clara	Floresta
Floresta_Brilho	Floresta
Mangue_Escuro	Mangue
Mangue_Clara	Mangue
Pastagem_Escura	Vegetação Herbácea
Pastagem_Clara	Vegetação Herbácea
Vegetação_Arbustiva	Vegetação Arbustiva
Restinga	Floresta
Urbano_Denso	Urbano
Urbano_Rarefeito	Urbano
Solo	Urbano
Água	Água

Em seguida foi utilizado o minerador de dados Weka 3.2 onde foi gerada uma árvore de decisão indicando os melhores descritores e limiares para a classificação de cada classe.

A validação do mapeamento foi feita a partir da geração de pontos aleatórios sobre o mapeamento (total de 180 pontos sendo 30 pontos para cada classe) através do software Arcgis 10, que em seguida foram analisados com auxílio do Google Earth, figura 3, gerando-se, em seguida, uma matriz de confusão.



Figura 3. Geração de pontos aleatórios sobre imagem Google Earth.

Considera-se que o uso do Google Earth como referência para validação pode ser considerado satisfatório pelo fato da escala alvo ser 1:100.000. Como este programa da Google permite a visualização de imagens de alta resolução para a área, é possível perceber detalhes suficientes e observar áreas que seriam inacessíveis mesmo em visitas a campo.

A seguir tem-se o fluxograma com o resumo das etapas da metodologia desenvolvida no presente trabalho.

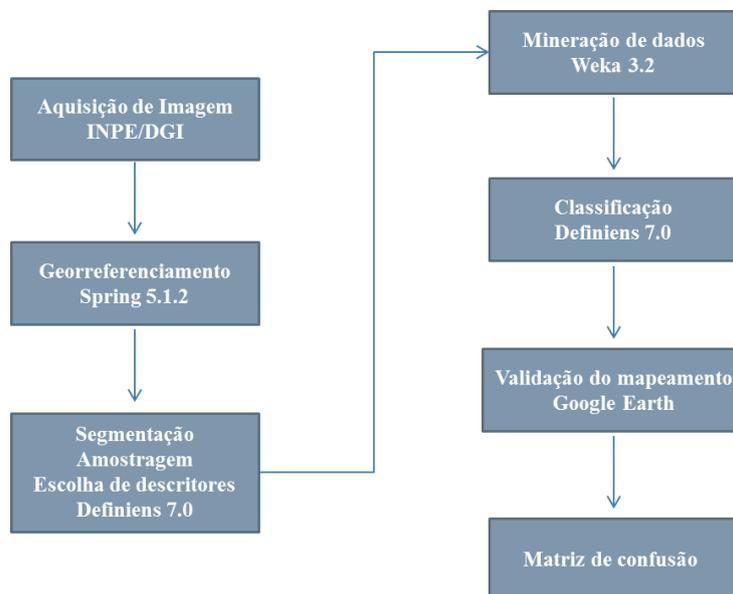


Figura 4. Fluxograma da Metodologia.

3. Resultados e Discussão

O primeiro resultado do presente trabalho foi uma árvore de decisão gerada pelo minerador de dados Weka 3.2 (Figura 5).

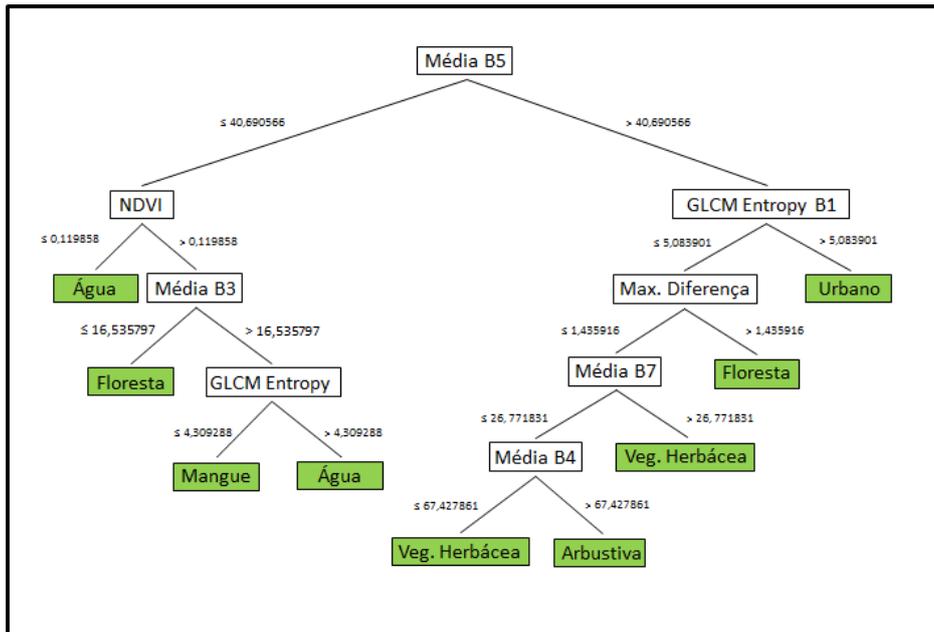


Figura 5. Árvore de decisão gerada pelo software Weka 3.2

A partir da mineração de dados foram escolhidos os melhores descritores (NÓS) e os intervalos para a classificação de cada classe em cada descritor. A escolha dos descritores, apresentados na tabela 2, utilizados nessa primeira etapa se baseou na utilização recorrente desses, em outros trabalhos de mapeamentos de cobertura e uso da terra.

Tabela 2. Lista de descritores selecionados para a Mineração de Dados.

Atributos/Descritores
Brilho
GLCM Entropy
GLCM Entropy B1
GLCM Entropy B2
GLCM Entropy B3
GLCM Entropy B4
GLCM Entropy B5
GLCM Entropy B7
Máxima Diferença
Mean B1
Mean B2
Mean B3
Mean B4
Mean B5
Mean B7
NDVI
Standard Deviation B1
Standard Deviation B2
Standard Deviation B3
Standard Deviation B4
Standard Deviation B5
Standard Deviation B7

A partir da árvore de decisão descrita anteriormente, foi possível gerar o mapa de vegetação a seguir, como pode ser visto na figura 6. É importante destacar que a classificação em questão não passou por nenhum tipo de edição manual ou alteração de modelo proposto pela árvore de decisão gerada no Weka.

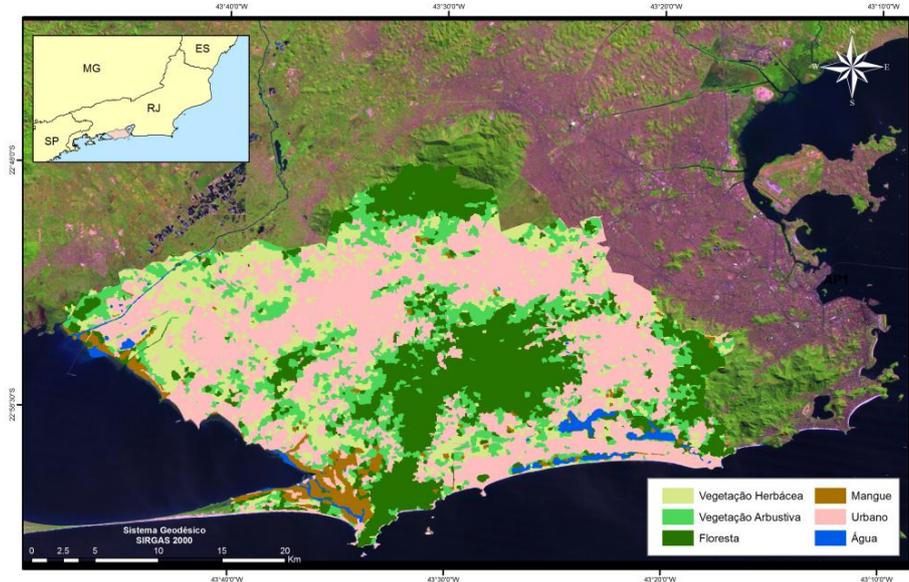


Figura 6. Mapa de vegetação da zona oeste da cidade do Rio de Janeiro.

Para avaliar a qualidade do mapeamento foi elaborada uma matriz de confusão (Tabela 3) onde foram analisados a exatidão global (Equação 1) e o grau de acerto por classe (Equação 2), dadas, respectivamente, pelas equações:

$$\frac{\sum \text{Diagonal de acerto}}{\text{Total de amostras}} \times 100 = \text{Exatidão Global} \quad (1)$$

$$\frac{\sum \text{Acerto por cada classe}}{\text{Nº de amostras para cada classe}} \times 100 = \text{Grau de acerto por classe} \quad (2)$$

Tabela 3. Valores obtidos para cada classe na matriz de confusão.

	Água	Vegetação Arbustiva	Floresta	Vegetação Herbácea	Mangue	Urbano
Água	30	0	0	0	5	0
Vegetação Arbustiva	0	14	4	6	2	2
Floresta	0	10	24	3	1	0
Vegetação Herbácea	0	3	0	16	0	1
Mangue	0	0	2	0	22	0
Urbano	0	3	0	5	0	27
Total de Amostras	30	30	30	30	30	30

A matriz de confusão constatou que o mapeamento elaborado apresentou exatidão global de 73,88% o que, segundo Cruz *et al.*, 2007, é considerado como de alto grau de exatidão, principalmente, por não ter passado por nenhum tipo de edição manual ou alteração de modelo. É nesse ponto que a metodologia proposta se destaca visto que, o maior esforço e gasto de tempo na realização de um mapeamento de cobertura e uso se encontram justamente nas etapas de modelagem e edição.

Analisando a matriz de confusão de forma simples, foi possível observar que as maiores confusões ocorreram entre a vegetação arbustiva e a floresta. Essa confusão se deve, em parte, pela dificuldade na etapa de amostragem da classe vegetação arbustiva, por seu típico caráter de transição.

Complementada a análise de qualidade do mapeamento foi calculado o grau de acerto por classe, como mostra a figura 7, onde foi possível observar que os maiores erros se concentram nas classes de vegetação arbustiva e herbácea, o que pode ser explicado pela semelhança espectral dessas duas classes, principalmente no contexto de um mapeamento em mesoescala.

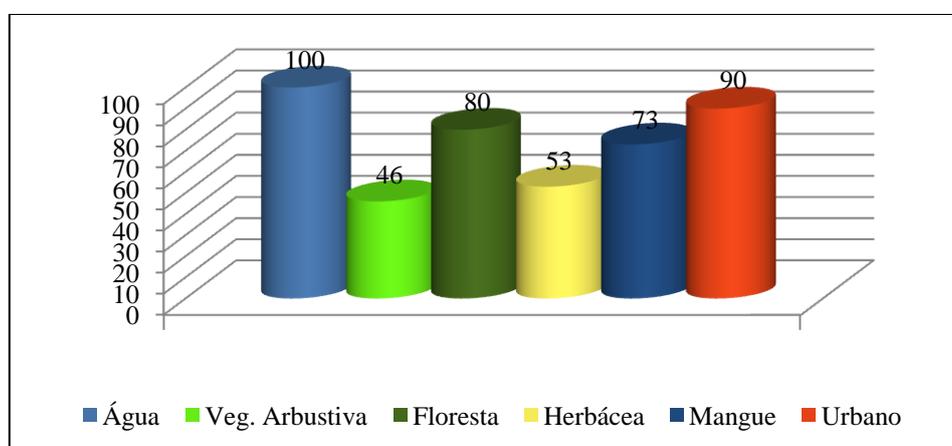


Figura 7. Exatidão por Classe de Mapeamento

4. Conclusões

A classificação a partir da metodologia proposta obteve resultados satisfatórios (73,88% de acerto), principalmente por não ter passado por nenhuma forma de edição manual ou alteração de modelo proposto no processo de mineração de dados.

Vale ressaltar que a utilização do minerador de dados, que além de possibilitar um grande ganho em tempo e esforço na etapa de modelagem, obteve resultados positivos na classificação de classes como, por exemplo, o mangue, conhecidas pela dificuldade da criação de modelos que não demandem tanto esforço na etapa de edição e ajustes.

Foi possível observar que a segmentação continua sendo um dos maiores desafios para a realização desse tipo de mapeamento, já que ainda se tem pouco conhecimento do melhor ajuste dos parâmetros de segmentação de acordo com a escala proposta e imagens utilizadas. A segmentação mesmo que tenha como objetivo a divisão da imagem em porções “homogêneas” (semelhante resposta espectral) ainda não atende completamente à identificação de objetos, que se caracterizam não somente pela resposta espectral, mas também pela forma, textura, extensão, vizinhança e proximidade. Dessa forma, ainda se faz necessário um maior esforço e conseqüente gasto de tempo na realização dessa etapa.

As dificuldades encontradas na amostragem, principalmente da classe da vegetação arbustiva pode ser solucionada a partir da realização de trabalhos de campo, para que não existam dúvidas quanto à pertinência de cada amostra.

A realização da pesquisa possibilitou concluir o ganho em tempo e esforço provenientes da utilização do minerador de dados que, apesar de ainda ser pouco explorado no contexto da classificação de imagens, se apresentou como uma ferramenta de grande ajuda para realização desse trabalho.

Referências Bibliográficas

CRUZ, C. B. M; VICENS, R. S.; SEABRA, V. S.; REIS, R. B; FABER, O. A.; RICHTER, M.; ARNAUT, P. K. E. & ARAÚJO, M. Classificação Orientada a Objetos no Mapeamento dos Remanescentes da Cobertura Vegetal do Bioma Mata Atlântica, Na Escala 1:250.000. In: XIII SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO Anais do XIII Simpósio Brasileiro de Sensoriamento Remoto, Florianópolis, 2007, INPE, p. 5691- 5698.

NOVO, E. M. L. M. - Sensoriamento Remoto, Princípios e Aplicações – Edgard Blucher, 1998.