

Classificação da Cobertura da Terra, Utilizando os Programas Livres: InterIMAGE, WEKA e QuantumGIS

Anderson Félix do Nascimento¹
Igor Brum Rubim²
Eduardo Gustavo Soares Pereira³
Rafael Silva de Barros¹
Monika Richter³

¹ Universidade Federal do Rio de Janeiro - UFRJ/IGEO
Rua Athos da Silveira Ramos, 274 – Bl. I – Sala 012, CCMN, Cidade Universitária – CEP
21941-590 – RJ, Brasil
{andinhofelix, rafael.barros}@ufrj.br

² Instituto Nacional de Colonização e Reforma Agrária - INCRA
Serviço de Meio Ambiente – Rio de Janeiro – RJ, Brasil
igor.brum@rjo.incra.gov.br

³ Universidade Federal Rural do Rio de Janeiro – UFRRJ/IM
Avenida Governador Roberto Silveira, Aero Clube – Nova Iguaçu – RJ, Brasil
edu_vcd@hotmail.com, mrichter84@hotmail.com

Abstract: Working with geomatics, remote sensing and its products like high resolution satellite images offers great data amount. Investigate and study those data demands appropriate software tools. This paper intended to present an image analysis for remote sensing based only in open source and free software. To reach this goal InterIMAGE 1.28, QuantumGIS 1.8 and WEKA 3.6 were employed. A GeoEye-1 satellite image acquired over Angra dos Reis city, Rio de Janeiro, Brazil, has been segmented in InterIMAGE. The segmentation file was used to collect some samples and attributes were exported with them on a vector file. A data mining software (WEKA) was used to analyse those data exported and a decision tree has been generated by the same software. Its results with decision rules were applied on the segmentation file in QuantumGIS. In this software, the advanced search tool has been used to find the group of features of their classes, according the decision rules and attributes established in WEKA. No edition or modification on the modeling proposed by WEKA was done. As result, a map with land cover and land use of the area was produced. A point vector file, with 30 samples per class was produced in QuantumGIS in order to evaluate accuracy of this map. As reference it was used an existing and validated land cover and land use map in 1:5.000 scale, produced by ESPAÇO laboratory using Definiens. With this comparison a confusion matrix has been generated, obtaining 66% of global accuracy.

Palavras-chave: Remote Sensing, Object Based Classification, Data Mining, InterIMAGE, Geoeye, Sensoriamento Remoto, Classificação Baseada em Objeto, Mineração de Dados.

1. Introdução

O sensoriamento remoto e os produtos gerados por satélites com sensores de alta resolução fornecem dados com excelente quantidade. Com a análise desses dados é possível produzir informações para classificação de cobertura e uso da terra. Através da análise das imagens podem ser gerados mapas temáticos que possuem classes, tais como: vegetação arbórea, vegetação rasteira, solo exposto, área urbana, vias urbanas, entre outras. Os mapas temáticos são extremamente úteis para gerenciamento, manejo, controle, cadastramento e acompanhamento da dinâmica e comportamento dos objetos inerentes às áreas estudadas.

A análise desses dados depende de programas (*softwares*) com ferramentas adequadas. As análises de imagens de satélite têm sido feitas com algoritmos que processam os dados "pixel-a-pixel" ou por regiões, através de algoritmos de segmentação. Essas técnicas normalmente produzem resultados que precisam de grande intervenção humana para se obter melhor qualidade. Análises baseadas em conhecimento buscam aproximar a classificação automática

da classificação feita pelo homem. A vantagem dessas novas técnicas de interpretação automática de imagens é o aumento da produtividade e precisão do trabalho, bem como na redução da subjetividade do resultado (Camargo *et al.*, 2012).

A utilização de *softwares* proprietários onera os trabalhos de profissionais que realizam análise de imagens de satélite e impede o acesso ao código fonte do programa (Steiniger e Hay, 2009). Alguns programas livres foram desenvolvidos e vêm sendo testados junto aos desenvolvedores. Programas livres que estão no mercado há alguns anos já são utilizados com qualidade equiparada aos programas de licenças pagas aos proprietários. Como exemplos, podemos citar o InterIMAGE e o SPRING .

Especificamente, a Análise Orientada a Objeto para interpretação de imagens de satélite encontra no InterIMAGE a solução livre para realizar as investigações. O InterIMAGE é um programa que permite análise de dados orientada a objeto, através da interpretação automática de imagens (Laboratório de Visão Computacional, 2012). É um produto da cooperação do Laboratório de Visão Computacional, do Departamento de Engenharia Elétrica da Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio) e das divisões de Processamento de Imagens e de Sensoriamento Remoto do Instituto Nacional de Pesquisas Espaciais (INPE) (Laboratório de Visão Computacional, 2012).

1.1. Objetivo

O objetivo deste trabalho foi utilizar programas livres e gratuitos para classificar a cobertura e uso da terra, utilizando técnicas de mineração de dados e análise orientada a objeto. Para isto, os programas InterIMAGE e WEKA e foram utilizados respectivamente para realizar a segmentação de imagem de satélite de uma parte da região central do município de Angra dos Reis - RJ e para mineração de dados. Para simplificar o processo de inserção dos dados minerados pelo WEKA no InterIMAGE, o mapeamento da cobertura e uso da terra o programa Quantum GIS (QGIS) foi empregado.

1.2. Área de Estudo

Como o software InterIMAGE ainda apresenta limitações quanto a extensão do arquivo imagem e, neste trabalho, o principal objetivo era testar a metodologia, selecionou-se uma pequena área de 500m por 500m (1.000x1.000 pixels) referente ao centro do município de Angra dos Reis no litoral sul do Rio de Janeiro. Nesta área as altitudes variam entre 0 e 150 metros.

Esta área foi escolhida por apresentar 2 tipos diferentes de vegetação e diferentes tipos de construções. Além disso, o Laboratório ESPAÇO de Sensoriamento Remoto e Estudos Ambientais, da geografia da UFRJ, já havia gerado um mapeamento em escala cadastral (1:5.000), seguindo metodologia consagrada e com rigorosa validação, com extensos levantamentos de campo, sobrevoo com helicóptero no âmbito de um projeto de mapeamento de áreas de risco no município com relação à deslizamentos de encostas.

2. Materiais e métodos

Os procedimentos metodológicos aplicados são esquematizados na figura 1 que apresenta os materiais utilizados e a classificação realizada através dos operadores e regras de decisão posteriormente validada em campo para avaliação e conclusão do mapa de cobertura da terra.

Como se pode ver na figura 1, foi utilizada uma imagem do satélite GeoEye-1, já ortorretificada, com as bandas 1, 2, 3 e 4 fusionadas, com sistema de projeção UTM, fuso 23 e sistema geodésico SIRGAS2000 (figura 2). O sistema InterIMAGE (versão 1.28), o sistema WEKA (versão 3.6) e o sistema QuantumGIS (versão 1.8.0) foram utilizados para segmentação, amostragem e geração dos atributos (descritores); mineração de dados e classificação; mapeamento, respectivamente.

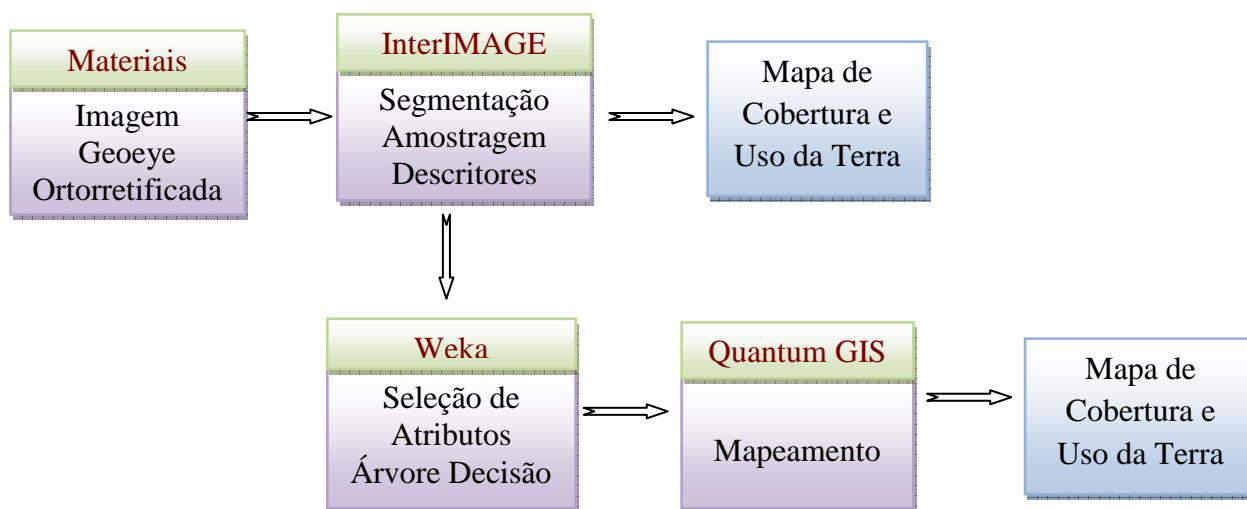


Figura 1: Resumo das etapas desenvolvidas

2.1 Segmentação, amostragem e geração dos atributos

A imagem fusionada foi inserida no Sistema InterIMAGE, onde se realizou a segmentação através da ferramenta *sample editor*. O operador "Crescimento de regiões" (*TA_Region_Growing_Segmenter*) foi escolhido com os seguintes parâmetros: Distância euclidiana igual a 20, Polígonos mínimos por área igual 12, *Use optimization - Yes* e *Reliability* igual a 0.2. Foram gerados 21.934 segmentos. Após a segmentação, procedeu-se à coleta de amostras para as classes estipuladas. As classes com as respectivas quantidades de amostras foram as seguintes: vegetação arbórea e arbustiva (37), vegetação rasteira (14), telhados cerâmicos (59), telhados metálicos (38), sombras em área urbana (15), sombras de vegetação (18), solo exposto (17), tetos e áreas com cimento (17), asfalto (17). O local das amostras está visível na figura 2. Em seguida foi realizada a exportação dessas amostras, com os seguintes atributos: brilho, entropia da banda 1, entropia da banda 2, entropia da banda 3, entropia da banda 4, média da banda 1, média da banda 2, média da banda 3, média da banda 4, média da banda 4 mais a média da banda 1, média da banda 4 menos a média da banda 1, média da banda 4 dividida pela média da banda 1, média da banda 4 dividida pela média da banda 3. O arquivo resultante da segmentação contendo as amostras foi utilizado para classificar a imagem, através do operador *top down TA_C45_Classifier* para todas as classes, a partir da marcação *multiclass* na classe arbórea. Cabe destacar que, para se chegar à um resultado mais apurado, o InterIMAGE nos permite a escolha de um classificador diferente para cada classe e regras de decisão *bottom-up* que utilizam lógica Fuzzi (Novack, 2009). No entanto, isso demandaria um tempo muito maior para a verificação do classificador mais indicado para cada classe.

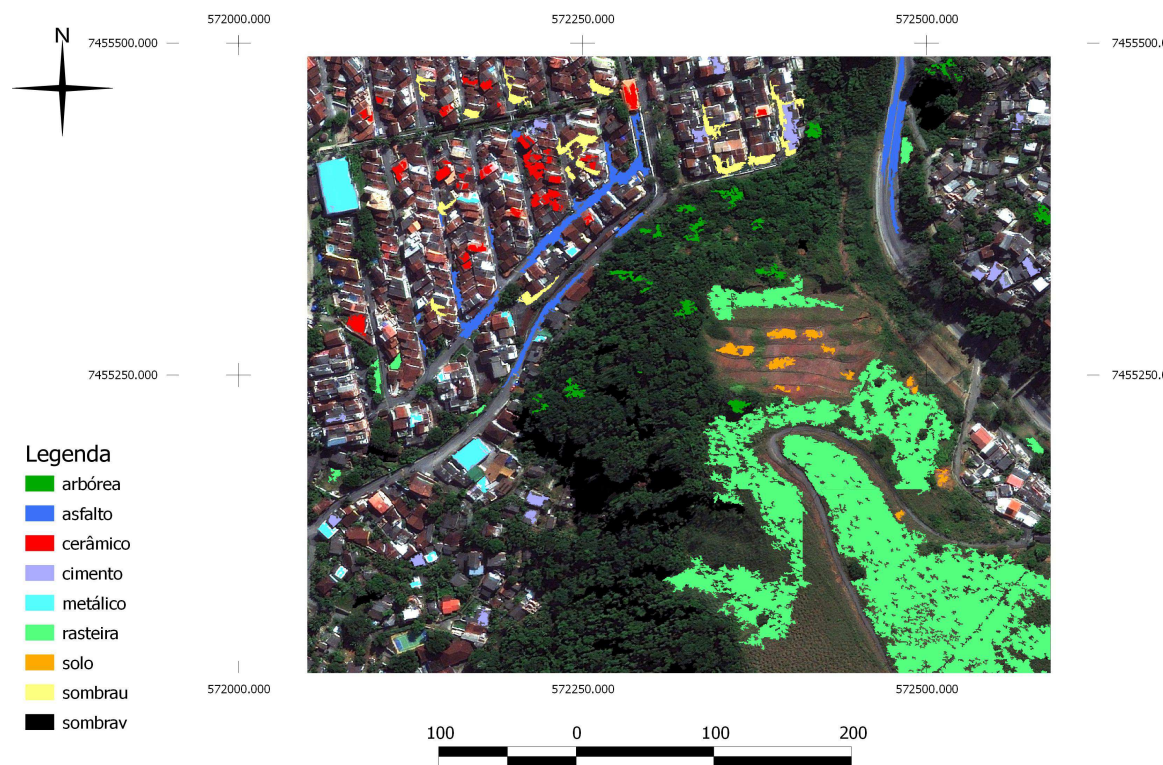


Figura 2: Imagem GeoEye-1 e amostras coletadas.

Foram gerados diversos descritores, dentre eles: brilho, média de cada banda, média de todas as bandas, entropia de cada banda, NDVI, razões entre bandas etc.

2.2 Mineração de dados

Após a geração dos atributos, foi feita a conversão da tabela de atributos do arquivo contendo as amostras que estavam no formato digital de banco de dados (.dbf) para o formato digital de "valor separado por vírgula" (.csv). Assim, este arquivo foi utilizado no programa WEKA para gerar a árvore de decisão. Através da ferramenta Explorer, o classificador utilizado foi *trees* "J48" para gerar a árvore em questão. Podemos observar que ao gerar a árvore o programa desconsidera alguns atributos que escolhemos, por serem insignificantes ou redundantes para a classificação. O WEKA também gera um matriz de erro (tabela 1), relacionada com as amostras coletadas, além de possuir algoritmos capazes de classificar e ranquear os melhores atributos ou os mais representativos para aquele conjunto de amostras. Cabe ressaltar ainda que, assim como o WEKA, o InterIMAGE, também, gera um arquivo com a árvore de classificação. Porém apenas no formato escrito e não gráfico.

Tabela 1: Matriz de erro gerada pelo programa WEKA 3.6.

```

==== Confusion Matrix (WEKA)====
 a b c d e f g h i <-- classified as
34 0 0 0 0 2 0 0 1 | a = arborea
 0 14 0 3 0 0 0 0 0 | b = asfalto
 0 1 46 2 2 0 8 0 0 | c = ceramico
 0 6 1 9 0 1 0 0 0 | d = cimento
 0 0 0 0 38 0 0 0 0 | e = metalico
 2 0 0 0 0 12 0 0 0 | f = rasteira
 0 0 6 0 0 0 11 0 0 | g = solo
    
```

0 0 0 0 0 0 0 15 0 | h = sombrau
1 0 0 0 0 0 0 0 17 | i = sombrav

2.3 Mapeamento

Com as regras de decisão construídas pela mineração de dados feita no WEKA, o arquivo contendo a segmentação com os atributos foi inserido no Quantum GIS para mapeamento das classes segundo os limiares para os atributos selecionados pelo WEKA. Abaixo estão apresentadas as etapas do mapeamento:

1º) Busca dos segmentos (feições) que satisfaziam as regras que definiam cada classe. As regras foram estipuladas de acordo com valores dos atributos que definiam as classes. Essa busca foi construída e efetuada com auxílio da ferramenta de "busca avançada" na tabela de atributos da camada referente ao arquivo alvo.

2º) Após encontrar os segmentos, selecionava-se as feições e a coluna referente ao nome da classe era preenchida.

3º) A operação era repetida mudando as regras de decisão para cada classe. Atributos e valores eram alterados de acordo com as especificações de cada classe.

4º) Algumas classes só faziam sentido durante a fase de amostragem e classificação, por apresentar respostas espectrais diferentes. Após a identificação de cada uma delas, elas foram reclassificadas:

- telhados cerâmico, metálico e de cimento e sombra em área urbana: reclassificados para Edificações;
- sombras de vegetação e arbórea e arbustiva: reclassificados para Arbórea e Arbustiva;

5º) As feições da mesma classe, mas que se encontravam em diferentes segmentos eram unidos em apenas uma linha da tabela de atributos, através de um comando de merge. Assim, ao final do processo de classificação o número de feições é igual ao número de classes.

2.4 Controle e verificação da verdade de campo

Após obter o mapa vetorial com as classes, necessita-se conferir o resultado comparando com informações que melhor traduzem a realidade. Para tanto, no Quantum GIS, a ferramenta de investigação vetorial "Pontos Aleatórios" foi empregada para gerar trinta (30) amostras de cada classe. Esse arquivo de pontos foi comparado com um mapeamento 1:5.000 da mesma área, feito anteriormente usando o software Definiens, sem uso de mineração de dados e validado com extensos levantamentos de campo. Como o mapa de referência apresentava mais classes do que o gerado neste trabalho, foi necessário compatibilizar as mesmas antes da validação. A comparação consistiu em criar um coluna de resultados no arquivo de pontos das amostras e relacionar ocorrências de igualdade ou diferença entre as amostras e o mapa de referência. Após essa comparação, foi montada a matriz de confusão.

3. Resultados

A metodologia utilizada na geração do mapeamento da cobertura por meio da integração das plataformas InterIMAGE, WEKA e QuantumGIS obteve bons resultados, simplificando e acelerando os procedimentos e disponibilizando uma metodologia totalmente apoiada em softwares livres.

O software InterIMAGE é uma importante alternativa ao uso do Definiens/e-Cognition nas classificações baseadas em objetos. Porém, ainda apresenta limitações que precisam ser superadas para que seu uso seja mais disseminado.

A limitação no tamanho das imagens a serem inseridas no sistema é uma delas. Tanto em mapeamentos em grande escala, que demandam imagens de alta resolução espacial, quanto nos mapeamentos de mesoescala, de usam imagens de média resolução, porém cobrem áreas

de maiores extensões, costuma ser necessário o uso de imagens que extrapolam os limites atuais do sistema.

A segmentação e o processo de amostragem e geração dos descritores são bastante satisfatórios. O arquivo de segmentação exportado para o formato shapefile, porém, apresentou problemas relacionados a topologia e precisou ser corrigido no ArcGIS, pois não encontramos solução no QuantumGIS. Esta correção só foi necessária pois optamos por gerar o arquivo de pontos de avaliação de forma aleatória, com um número pré-determinado de amostras em cada classe.

Em diversos momentos durante o uso do InterIMAGE o sistema falhou/fechou, sendo necessário repetir alguns passos já executados mas não salvos antes do encerramento inesperado.

O software WEKA mostrou-se uma excelente alternativa à etapa de modelagem, que normalmente é custosa em termos de tempo, além de ser possível – senão provável – que não se faça a melhor seleção de atributos e/ou determinação de limiares. O uso do WEKA é simples, exigindo apenas a organização e formatação do arquivo de entrada segundo seu padrão. Na versão mais nova do sistema, já é possível utilizar o arquivo de entrada com a extensão usada pela planilha eletrônica mais difundida no mercado. A figura 3 apresenta a árvore de decisão gerada pelo WEKA, com os atributos selecionados e limiares indicados.

Como o WEKA seleciona não apenas os melhores atributos, mas também indica os limiares de separação entre as classes, optou-se por implementar o mapeamento no QuantumGIS, evitando-se o retorno ao InterIMAGE e a necessidade de respeito ao seu modelo conceitual. Assim, por consulta/seleção em banco de dados, utilizou-se os limiares nos atributos indicados para selecionar os segmentos que pertencem a cada classe. O resultado obtido pode ser observado na figura 4.

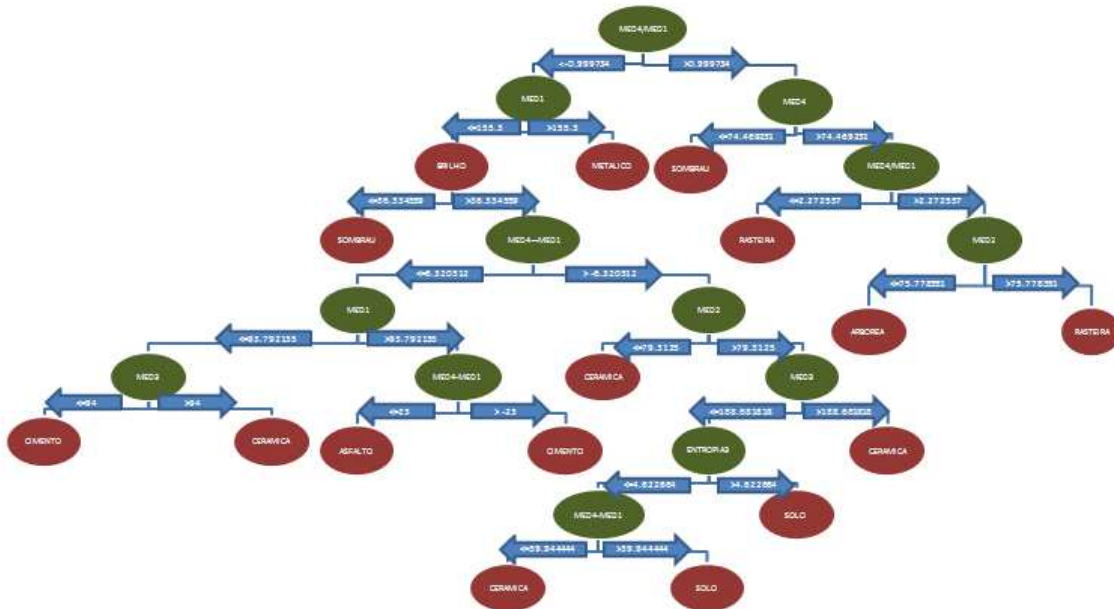


Figura 3: Árvore de decisão do WEKA

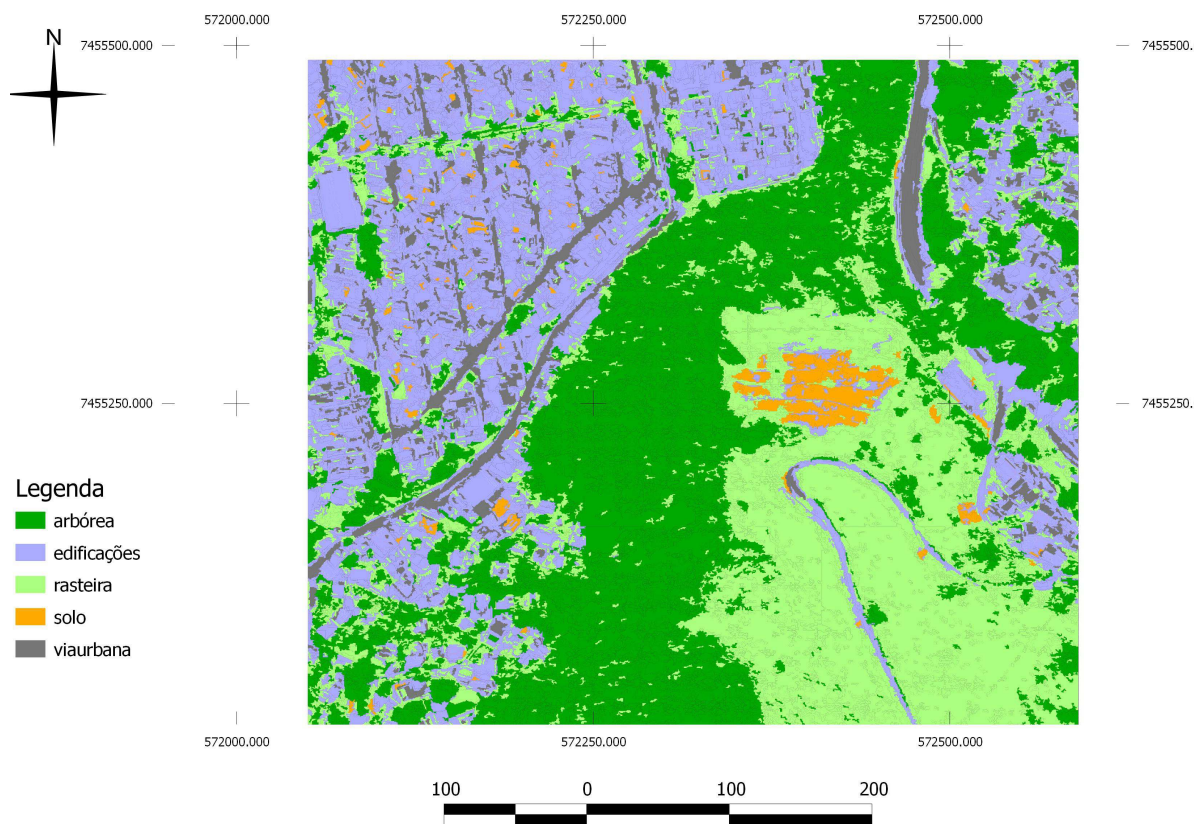


Figura 4: Classificação final no QuantumGIS

A partir da comparação entre os resultados obtidos pela classificação no QuantumGIS e a realidade no terreno (representada pelo mapa de cobertura e uso da terra, na escala 1:5.000, gerado anteriormente pelo laboratório ESPAÇO), foi elaborada a matriz de confusão (ver tabela 2). Ao aplicar o método da exatidão global, que consiste em verificar a razão entre os pontos classificados corretamente e total de pontos, chegamos a um percentual de 66% de exatidão. Entre os percentuais de acerto por classe, a que se verificou o maior percentual foi a classe edificações (83%) e a que obteve o menor percentual foi a classe Vias (36%). Diante do bom resultado alcançado, conclui-se que a integração dessas plataformas livres e gratuitas deve ser mais explorada em trabalhos futuros.

Tabela 2: Matriz de confusão da classificação final com a realidade do terreno

Realidade do terreno \ Classificação	Classificação				
	Árborea	Rasteira	Solo	Edificações	Vias
Árborea	21	4	0	0	1
Rasteira	4	21	2	0	0
Solo	0	0	21	1	0
Edificações	5	5	7	25	18
Vias	0	0	0	4	11

Exatidão por Classe	70%	70%	70%	83%	36%
Exatidão Global (EG)	66%				

Considera-se que a Exatidão Global de 66% indica um bom resultado pelo fato de não ter sido feito nenhum ajuste na modelagem gerada pelo WEKA (aceitou-se todos os limiares e atributos selecionados) e nenhuma edição no mapa gerado. Analisando-se cada classe individualmente, observa-se um bom desempenho, exceto para a classe Vias Urbanas, que se confundiu muito com as Edificações. A segmentação e a amostragem podem ser revistas para melhorar o percentual de acerto, mas acredita-se que seria necessário, também, um ajuste no modelo proposto pelo WEKA para separação desta classe.

4. Conclusões e Considerações Finais

De forma geral, este trabalho atendeu seus objetivos, tanto com relação a gerar o mapa de cobertura e uso da terra fazendo uso da classificação baseada em objetos quanto na expectativa de se conseguir fazê-lo usando apenas sistemas livres. O teste do uso do minerador WEKA como único recurso para modelagem também se mostrou promissor, visto que esse obteve uma boa exatidão Global e permite uma redução muito grande no tempo de modelagem. É preciso considerar, também, que o acerto elevado foi favorecido pelo reduzido número de classes no mapeamento final.

Apesar de dispor da maioria das ferramentas necessárias para execução do trabalho, o software InterIMAGE ainda necessita de um maior amadurecimento, especialmente no que se refere ao tamanho máximo do arquivo a ser trabalhado e na estabilidade do sistema, que ainda interrompe inesperadamente muitas vezes.

Outro aspecto a ser destacado refere-se a metodologia proposta, com o uso de um sistema de banco de dados geográficos – neste caso foi usado o QuatumGIS – para implementação da modelagem gerada pelo Weka. Como o minerador seleciona os atributos que melhor diferenciam as classes indicadas com valores de limiares para cada um deles fica fácil selecioná-los no Shapefile com os valores dos atributos em todos os segmentos. Isso simplifica bastante o mapeamento, pois é uma tarefa mais simples do que implementar a modelagem no InterIMAGE.

Bibliografia:

Laboratório de Visão Computacional. Pontífice Universidade Católica do Rio de Janeiro (2012). InterIMAGE1.30 **Manual do Usuário**. Disponível em: < <http://www.lvc.ele.puc-rio.br/projects/interimage/pt-br/documentacao/arquivos/InterIMAGE%201.30%20-%20Manual%20do%20Usu%C3%A1rio.pdf>> Acesso em: 14.nov.2012.

Steiniger, S.; Hay, G. J. Free and open source geographic information tools for landscape ecology. **Ecological Informatics**, v. 4, n. 4, p. 183–195, 2009.

Camargo, F.F.; Almeida, C.M.; Costa, G.A.O.P.; Feitosa, R.Q.; Oliveira, D.A.B.; Heipke, C.; Ferreira, R.S. **An open source object-based framework to extract landform classes**, v. 39, n. 1, p. 541-554, 2012.

Novack, T. **Classificação da cobertura da terra e do uso do solo urbano utilizando o sistema InterIMAGE e imagens do sensor Quickbird**. 2009. 214 p. (INPE-14183-TDI/1095). Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos.