

USO DE MINERAÇÃO DE DADOS PARA MAPEAMENTO DA COBERTURA DA TERRA EM IMAGEM OLI/LANDSAT 8 NO PARQUE NACIONAL DO ITATIAIA

Leonardo Herms Maia *

Gustavo Mota de Sousa **

Manoel do Couto Fernandes ***

Resumo: O objetivo deste trabalho é a aplicação das técnicas de mineração de dados para o mapeamento da cobertura da terra Parque Nacional do Itatiaia usando o software InterIMAGE e imagens do Sensor OLI / Landsat 8. A área de estudo abrange parte do Estado de Minas Gerais e Rio de Janeiro com destaque a região do Sul de Minas e o Planalto das Agulhas Negras que possuem remanescentes do bioma Mata Atlântica. Os dados utilizados foram a banda pancromática e as bandas multiespectrais (2-7) do sensor OLI / Landsat 8 que realizaram a geração de objetos por segmentação em Baatz e a extração de atributos, respectivamente. As classes são definidas para o mapeamento de áreas urbanas, gramínea, floresta, rocha, sombra e água. Foram coletadas amostras no destinadas a mineração de dados através do Classificador C4.5. O resultado observado a partir da validação do mapeamento indica a exatidão global de 72% e um índice kappa de 0,66, demonstrando resultados promissores e a eficácia da metodologia aplicada.

Palavras-chave: Mapeamento de cobertura, Landsat 8, Mineração de Dados, InterIMAGE.

DATA MINING FOR LAND COVER MAPPING WITH OLI/LANDSAT 8 IN ITATIAIA NATIONAL PARK

Abstract: The objective of this work is to apply data mining technique for land cover mapping of Itatiaia National Park using InterIMAGE software and images of OLI/Landsat 8 sensor. The study area is part of the Atlantic Forest and covers Minas Gerais and Rio de Janeiro states especially the Minas southern region and the Agulhas Negras Highlands. The data used were the panchromatic band and the multispectral bands (2-7) of the OLI / Landsat 8 that performed the generation of objects with Baatz segmenter and the extraction of attributes, respectively. The classes are defined for mapping urban areas, grass, forest, rocky, shadow and water that were collected by samples editor aimed at data

* Geógrafo pela UFRRJ - lhmaia.geo@gmail.com

** Professor Adjunto, Dep. de Geociências/UFRRJ – gustavoms@ufrj.br

*** Professor Associado, Dep. de Geografia/UFRRJ – manoel.fernandes@ufrj.br

mining through the algorithm C4.5 Classifier. The result observed from the mapping validation indicates global accuracy of 72% and a kappa index of 0.66, demonstrating promising results which show the effectiveness of the applied methodology.

Keywords: Land cover, Landsat 8, Data Mining, InterIMAGE.

USO DEL MINERÍA DE DATOS PARA EL MAPEO DE COBERTURA DE LA TIERRA EN IMAGEN OLI/LANDSAT 8 DEL PARQUE NACIONAL ITATIAIA

Resumen: El objetivo del trabajo es la aplicación de técnicas de minería de datos para el levantamiento cartográfico de la cobertura de tierra del Parque Nacional do Itatiaia utilizando el software InterImage e imágenes del Sensor OLI / Landsat 8. El área de estudio abarca parte de la provincia de Minas Gerais y Rio de Janeiro, destacándose la región sur de Minas y el Planalto das Agulhas Negras que poseen remanecientes del bioma Mata Atlántica. Los datos utilizados fueron las bandas pancromáticas y multiespectrais (2-7) del sensor OLI / Landsat 8 que generaron objetos por segmentación en Baatz y la extracción de atributos, respectivamente. Las clases están definidas para el levantamiento cartográfico de áreas urbanas, gramínea, floresta, rocha, sombra y agua. Fueron colectadas muestras no destinadas a la minería de datos a través del Clasificador C4.5. El resultado observado a partir de la validación del levantamiento cartográfico indica la exactitud global del 72% y un índice kappa de 0,66, demostrando resultados promisoros y la eficacia metodológica aplicada.

Palabras clave: Minería de datos, Cobertura de tierra, Landsat 8, InterIMAGE.

Introdução

A gestão das Unidades de Conservação necessita de planejamento para os diversos desafios que ocorrem dentro e fora dos seus limites e para isso é importante a elaboração de mapeamentos de cobertura da terra que indiquem uma representação da realidade identificada por meio de levantamentos aéreos e orbitais.

No que tange a realização de mapeamentos de cobertura da terra, o Sensoriamento Remoto é primordial e contribui decisivamente na interpretação da estrutura horizontal de uma paisagem (Sousa et al., 2012). Isso decorre de que o mapeamento da cobertura da terra é entendido como a caracterização dos objetos da superfície terrestre em termos de suas propriedades biofísicas, físicas e químicas que exibem

uma relação própria de interação energia-matéria, ou seja, um padrão mais ou menos definido de reflectância (Barnsley et al..., 2001; Prado, 2009). Em face ao exposto não se propõe o mapeamento do uso do solo, que de acordo com Novo (1992), refere-se à utilização cultural, ou a forma como o espaço está sendo ocupado pelo homem, como define Rosa (2009), e sim identificar a estrutura dos objetos na paisagem em análise, ou seja, sua distribuição espacial.

Nos últimos anos, o Sensoriamento Remoto tem se desenvolvido em diversas áreas que se destacam a aquisição de imagens orbitais com melhores resoluções espacial e espectral além de novos métodos de análise que têm sido aprimorados para suportar a geração desses novos bancos de dados. Esses avanços apontam para o desenvolvimento da GEOBIA (GEOgraphic Object-Based Image Analysis) indicado por Blaschke et al. (2014) como um novo paradigma para a análise de imagens aplicável à diversas áreas ambientais que envolvem a descoberta e a modelagem do conhecimento para mapeamentos com diversos fins. A descoberta do conhecimento tem como etapa principal a mineração de dados devido a realização da seleção dos métodos a serem utilizados para identificar padrões existentes em banco de dados, os quais são analisados para a busca por padrões de interesse que visam a identificação de fenômenos através dos melhores parâmetros que resolvam o objetivo em questão (Sousa, 2013).

A área de estudo abrange parte do Estado de Minas Gerais e Rio de Janeiro com destaque a região do Sul de Minas e o Planalto das Agulhas Negras que possuem áreas do bioma Mata Atlântica protegidas pelo mosaico de Unidades de Conservação da Serra da Mantiqueira em que se destaca do Parque Nacional do Itatiaia (Figura 1).

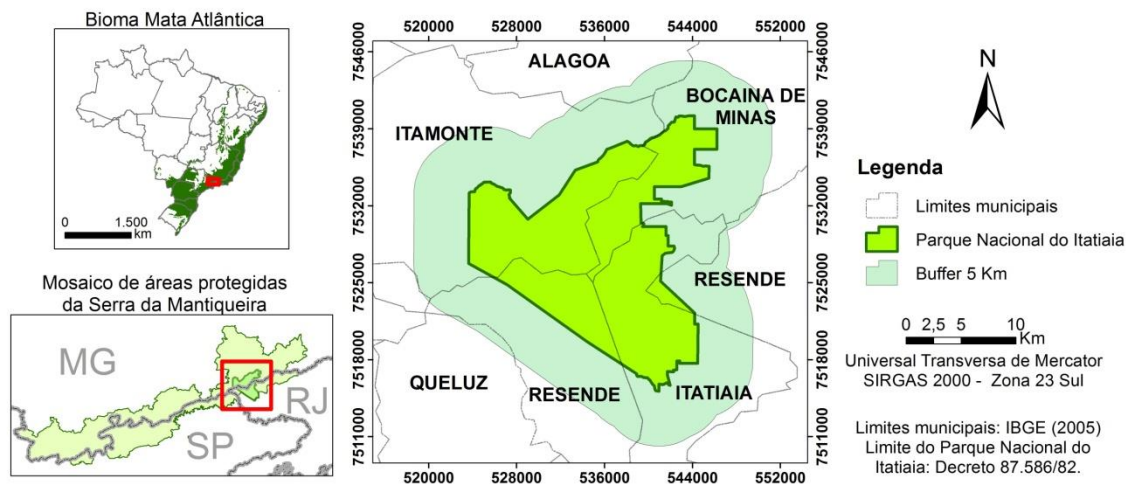


Figura 1. Localização do Parque Nacional do Itatiaia (Adaptado de Tomzhinski, 2012).

A proposta desse estudo é aplicar a técnica de mineração de dados para o mapeamento da cobertura da terra do Parque Nacional do Itatiaia e arredores por meio de uma imagem do sensor OLI/LANDSAT 8 e o software livre InterIMAGE. O projeto então se dividiu em quatro etapas: Aquisição da imagem LANDSAT 8 pelo site Earth Explorer, correção atmosférica com o modelo AtmCorOLI (Antunes et al., 2012), segmentação (Batz Segmenter) e coleta de amostras (Samples Editor), mineração de dados (C4.5_classifier) e validação do mapeamento de cobertura da terra.

Metodologia

A metodologia utilizada (Figura 2) teve como base a realização de técnicas de correção atmosférica de imagens, segmentação e mineração de dados resultando na geração do mapa de cobertura da terra.

A correção atmosférica foi realizada seguindo o modelo de transferência radiativa AtmCorOLI conforme descrito por Antunes et al.. (2012) e Sousa & Antunes (2013). Esse procedimento foi importante para corrigir os efeitos dos aerossóis presentes na atmosfera e que propiciam diferenças nas respostas espectrais nos alvos terrestres.

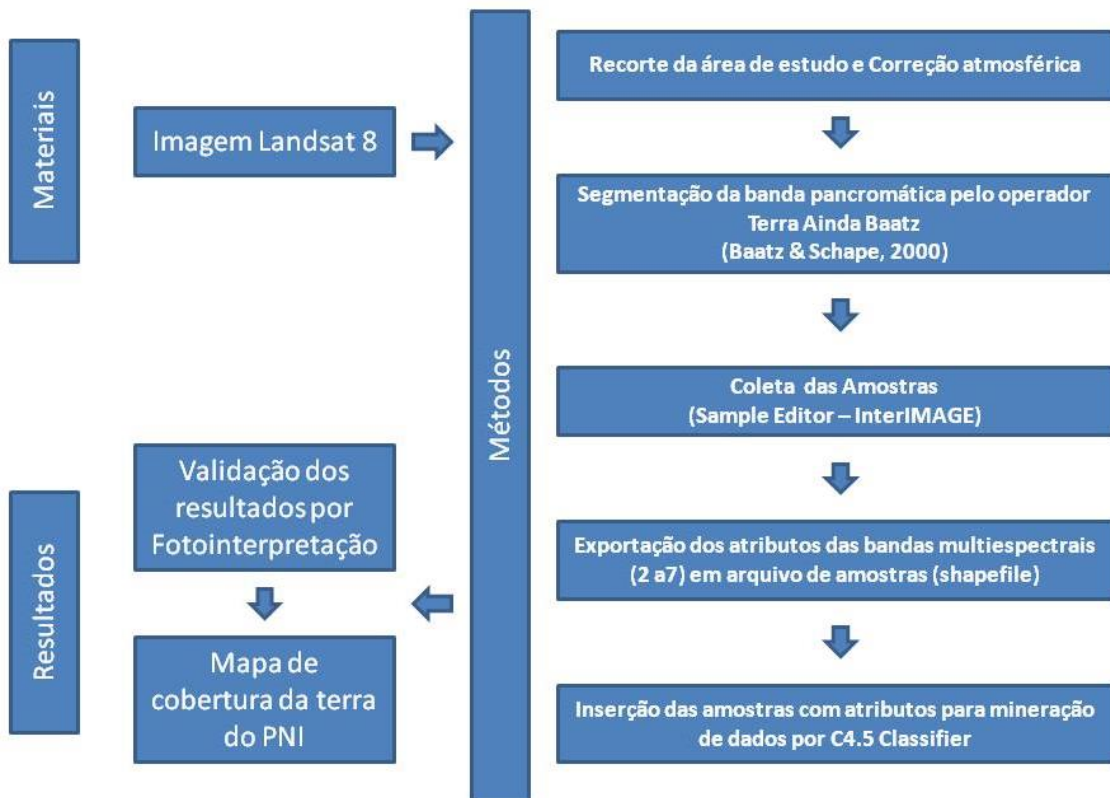


Figura 2. Fluxograma metodológico

Os procedimentos agora descritos foram realizados no software InterIMAGE (Costa et al., 2008) que tornou possível a geração da descoberta de conhecimento dos objetos com base na coleta de amostras e a mineração de dados para classificação da cobertura da terra.

A geração das amostras se deu a partir da segmentação da banda pancromática do LANDSAT 8 utilizando o operador "Terra Aída Baatz Segmenter" (Baatz & Schape, 2000) com os seguintes parâmetros: compacidade 0.8, cor 0.9 e escala 50. Os segmentos gerados foram analisados visando a delimitação das classes manualmente na imagem. Visando a aquisição de dados consistentes, foi realizada a coleta de 30 amostras de cada uma das classes definidas (área urbana, floresta, vegetação rasteira, sombra, rocha, água), de forma que o algoritmo pudesse distinguir as classes a serem classificadas. Após, as amostras foram exportadas junto com os atributos descritos na tabela 1.

Tabela 1: Lista dos Atributos (InterIMAGE, 2011)

Nome do atributo	Definição
Amplitude (amplitude value)	Representa a diferença entre os valores de pixel máximo e mínimo de um objeto para um segmento.
Entropia (entropy)	Medida estatística aleatória que pode ser usada para descrever algumas características de textura. Maior aleatoriedade dos dados leva a maiores valores de entropia.
Máximo valor do pixel (maxPixelValue)	O valor de pixel máximo encontrado dentro de um segmento.
Média (mean)	A média de valores de pixels dentro de um segmento.
Mínimo valor do pixel (minPixelValue)	O valor de pixel mínimo encontrado dentro de um segmento.
Moda (mode)	Representa o valor mais frequente entre um conjunto de valores.
Razão (ratio)	Representa a quantidade que a imagem contribui para o brilho total de um segmento.
Desvio Padrão (standardDeviation)	Representa o grau de dispersão numérica de dados em torno da média.
Soma dos pixels (sumPixelValue)	Representa a soma de todos os valores dos elementos dentro de uma área de um segmento.
Variância (variance)	Semelhante ao desvio padrão, a variância também representa o grau de dispersão de dados numéricos em torno da média, mas na escala original dos valores.

A coleta de amostras gerou um arquivo shapefile com os atributos descritos para cada banda espectral além de índices espectrais como o Índice de Vegetação da Diferença Normalizada (NDVI) que foi reintroduzido ao projeto do InterIMAGE com adaptação na rede semântica (Fig. 3A) para a inclusão do operador de mineração de dados baseado no algoritmo C4.5 (Quinlan, 1993). O operador C4.5 (Terra Aida C4.5 Classifier) gerou uma árvore de decisão (Fig. 3B) com base nas amostras inseridas e realizou a classificação da cobertura da terra.

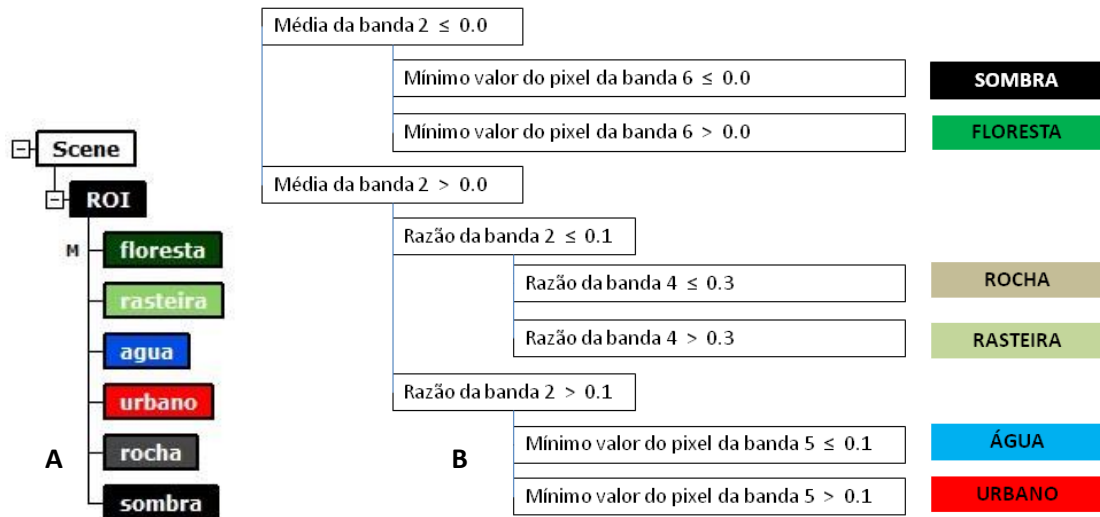


Figura 3. Rede semântica (A) e Árvore de decisão gerada pelo algoritmo C4.5 (B)

A árvore de decisão gerada por mineração de dados é bastante resumida e apresenta o uso das bandas 2 (azul), 4 (vermelho) e 5 (infravermelho próximo) e os atributos utilizados foram de média, razão e mínimo valor do pixel. Os resultados das classes de cobertura da terra foram confrontados com amostras de validação elaboradas com base nas imagens LANDSAT 8 e auxílio do Google Earth Pro para resolução de possíveis dúvidas.

Resultados

O mapa de cobertura da terra foi gerado através da metodologia realizada inteiramente no software InterIMAGE, posteriormente uma análise visual foi realizada objetivando-se verificar a coerência entre a imagem e os resultados alcançados. Caso fossem verificadas incongruências, novos procedimentos de coleta de amostras serão realizados para obtenção de classificação satisfatória.

A geração das amostras de validação se deu através da identificação por fotointerpretação de 30 amostras por classe bem distribuídas ao longo da área de estudo. As amostras de validação foram sobrepostas à classificação realizada e tornou possível à elaboração de uma matriz de confusão (Tabela 2), que fornece dados para o cálculo dos índices de Acurácia Global e Kappa (Congalton & Green, 1999).

Leonardo Herms Maia, Gustavo Mota de Sousa e Manoel do Couto Fernandes,
Uso de mineração de dados para mapeamento da cobertura da terra em imagem Landsat 8 no Parque Nacional do Itatiaia

Tabela 2. Matriz de confusão

Área - Teste		Classes geradas manualmente						
		Área Urbana	Floresta	Veg. Rasteira	Sombra	Rocha	Água	Total
Classes geradas automaticamente	Área Urbana	16	0	3	0	0	0	19
	Floresta	0	28	3	7	1	0	39
	Veg. Rasteira	6	9	32	0	13	0	60
	Sombra	0	0	0	25	0	0	25
	Rocha	0	0	3	0	13	0	16
	Água	0	0	6	0	0	17	23
	Total	22	37	47	32	27	17	182

O valor do índice global adquirido foi de 72%, enquanto o índice kappa atingiu o valor de 0.66 que, de acordo com Landis & Koch (1977), atingiram valores satisfatórios. Visando verificar o percentual de acertos por classe e sobreposições das classes na matriz de confusão, construiu-se uma matriz de confusão com valores percentuais apresentada na tabela 3.

Tabela 3. Matriz de percentual de acerto por classe

Área - Teste		Classes geradas manualmente					
		Área Urbana	Floresta	Veg. Rasteira	Sombra	Rocha	Água
Classes geradas automaticamente	Área Urbana	84,2	0	15,8	0	0	0
	Floresta	0	71,8	7,7	17,9	2,6	0
	Veg. Rasteira	10	15	53,3	0	21,7	0
	Sombra	0	0	0	100	0	0
	Rocha	0	0	18,8	0	81,3	0
	Água	0	0	26,1	0	0	73,9

Analisando a matriz de percentual (Tabela 3) obteve-se os resultados de acurácia do usuário: Sombra (100%), Área Urbana (84,2%), Rocha (81,3%), Água (73,9%), Floresta (71,8%), Vegetação Rasteira (53,3%).

A classe vegetação rasteira foi a que apresentou o menor valor percentual, isso é resultado da diversidade da distribuição com classes vizinhas e as características

apresentadas na agregação de diferentes classes de vegetação rasteira devido a variação altitudinal que atinge até 2300 m de altitude como é o caso das áreas com características de campo de altitude nas áreas de planalto (Colares et al., 2015). As áreas de vegetação rasteira demonstraram confusão com áreas de transição entre classes próximas em toda a área de estudo que são floresta (15%), rocha (21,7%) e áreas urbanas (10%).

As áreas de floresta também apresentaram confusão com áreas de transição com vegetação rasteira (7,7%), rocha (2,6%) e áreas sombreadas (17,9%). No caso das áreas identificadas como sombra, trata-se de uma classe que deve ser incorporada a outras classes por não ter características de cobertura da terra, sendo indicada para observação da geomorfologia local.

As classes de água, rocha e urbano se confundiram com o entorno das áreas de transição coberto com vegetação rasteira atingindo valores de 26,1%, 18,8% e 15,8% respectivamente.

O mapa de cobertura da terra do Parque Nacional de Itatiaia se apresenta na figura 4 e demonstra que a área possui grande diversidade de classes em que se destacam as áreas de vegetação rasteira e os fragmentos florestais que ocupam as áreas com maior declividade.

As áreas urbanizadas encontram-se principalmente ao longo da Rodovia Presidente Dutra (BR 116) e estão em crescimento devido a implantação de um parque industrial que dinamiza a cobertura da terra, ocupando áreas que antes eram de pastagem. Os afloramentos rochosos são delimitados na região do Planalto do Itatiaia e as áreas sombreadas merecem atenção na classificação por conter outras classes que poderão ser classificadas com base em outros parâmetros como a geomorfologia e proximidade.

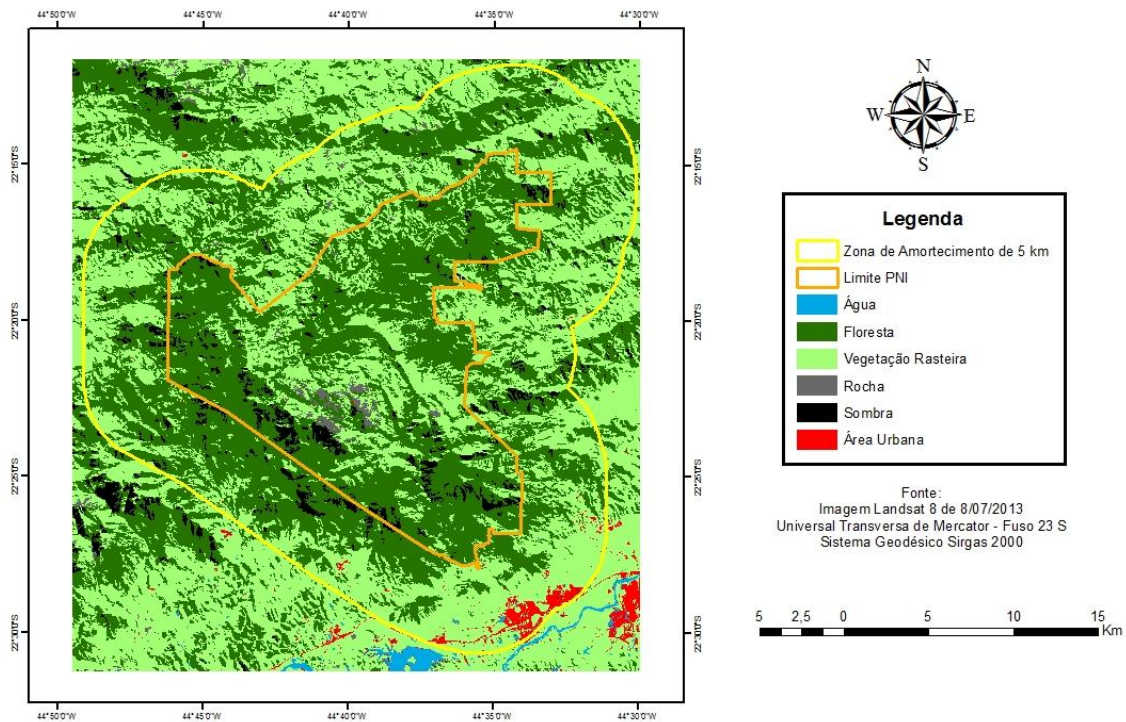


Figura 4. Mapa de cobertura da terra do Parque Nacional do Itatiaia

Conclusão

Os resultados de classificação mostram que o modelo baseado em mineração de dados em Imagem LANDSAT 8 utilizado no Parque Nacional do Itatiaia atingiu o objetivo, porém outros dados e parâmetros poderão ser inclusos e avaliados posteriormente visando o refinamento dos resultados, tais como modelo altimétrico, uso de índices espectrais de áreas urbanas e água, além de agregar a classe de sombra à outras classes ao seu entorno.

O InterIMAGE permitiu a utilização de bandas do sensor OLI do satélite LANDSAT 8 com diferentes resoluções espaciais com a segmentação baseada no operador Baatz Segmenter na banda pancromática (15 m) e, a extração de atributos das bandas multiespectrais (30 m) com a mineração de dados através do algoritmo C4.5. A extração dos atributos que geraram a árvore de decisão na rede semântica garantiu uma boa associação para a descoberta de conhecimento visando à construção de mapeamento para cobertura da terra.

Leonardo Herms Maia, Gustavo Mota de Sousa e Manoel do Couto Fernandes,
*Uso de mineração de dados para mapeamento da cobertura da terra
em imagem Landsat 8 no Parque Nacional do Itatiaia*

Os resultados dos índices Kappa e Acurácia Global demonstram que a metodologia foi eficaz e capaz de gerar um mapeamento que pode ser aprimorado utilizando novas fontes de dados e o potencial do sistema InterIMAGE com o emprego de outros operadores.

O InterIMAGE se encontra em desenvolvimento e ainda possui algumas limitações que vem sendo superadas, os resultados apontam ser uma alternativa eficaz, livre e gratuita, para análises de imagens baseadas em objeto.

Referências Bibliográficas

ANTUNES, M.A.H., DEBIASI, P., COSTA, A.R. DA.; GLERIANI, J.M. Correção atmosférica de imagens ALOS/AVNIR-2 utilizando o modelo 6S. *Revista Brasileira de Cartografia – RBC*, nº 64/Edição especial Sensoriamento Remoto: p. 531-539, 2012.

BAATZ, M., SCHÄPE, A., Multiresolution Segmentation – an optimization approach for high quality multi-scale image segmentation. In: Strobl, J. et al. (eds.): *Angewandte Geographische Informationsverarbeitung XII*. Wichmann, Heidelberg, p.p. 12-23, 2000.

BARNESLEY, M. J.; MOLLER-JENSEN, L.; BARR, S. L. Inferring urban land use by spatial and structural pattern recognition. In: Donnay, J.; Barnesley, M. J.; Longley, P. A. (Eds). *Remote sensing and urban analysis*. London: Taylor & Francis. cap. 7, p. 115-144., 2001.

BLASCHKE, T.; HAY, G.J.; KELLY, M.; LANG, S.; HOFMANN, P.; ADDINK, E.; FEITOSA, R.Q.; MEER, F.V.D.; WERFF, H.V.D.; COILLIE, F.D. TIEDE, D. Geographic Object-Based Image Analysis – Towards a new paradigm. In.: *ISPRS Journal of Photogrammetry and Remote Sensing*, V. 87, pp. 180–191, Jan. 2014.

COLARES, I.V.V.; NUNES, M.T.O; SOUSA, G.M.; FERNANDES, M.C. Análise do uso e cobertura do solo através da plataforma InterIMAGE aplicado ao Parque Nacional do Itatiaia – RJ. In.: *Revista Brasileira de Geomática – RBGeo*, nº 1, v.3, pp. 24-31, jan. 2015

CONGALTON, R. G., GREEN, K. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, Lewis Publishers, New York., 1999.

COSTA, G.A.O.P.; PINHO, C.M.D.; FEITOSA, R.Q.; ALMEIDA, C.M.; KUX, H.J.H.; FONSECA, L.M.G.; OLIVEIRA, D. (2008) InterIMAGE: Uma plataforma cognitiva open source

Leonardo Herms Maia, Gustavo Mota de Sousa e Manoel do Couto Fernandes,
Uso de mineração de dados para mapeamento da cobertura da terra em imagem Landsat 8 no Parque Nacional do Itatiaia

para a interpretação automática de imagens digitais. *Revista Brasileira de Cartografia – RBC*, nº 60/4, p.331-337, dez 2008.

InterIMAGE – Interpreting Images Freely, Disponível em <http://wiki.dpi.inpe.br/doku.php?id=interimage:operators_documentation>. Acesso em: 11 de nov. 2011.

LANDIS, J. R.; KOCH, G. G., 1977 The measurement of observer agreement for categorical data. *Biometrics*, v. 33, n. 1, p. 159-174.

NOVO, E. M. L. M. *Sensoriamento remoto: princípios e aplicações*. 3ª ed. revista e ampliada. São Paulo: Edgard Blucher, 388 p., 2008.

PRADO, F.A. *Sistema hierárquico de classificação para mapeamento da cobertura da terra nas escalas regional e urbana*. Dissertação (Mestrado). Programa de PósGraduação em Ciências Cartográficas: UNESP Presidente Prudente, 168 f., 2009.

ROSA, R. *Introdução ao Sensoriamento Remoto*. Uberlândia: EDUFU, 7ª Ed. 264 p., 2009.

QUINLAN, J. R. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.

SOUSA, G. M.; ANTUNES, M. A. H. Correção atmosférica de imagens AVNIR-2/ALOS com o modelo 6S no Parque Nacional do Itatiaia. In: Simpósio Brasileiro de Sensoriamento Remoto, 16. (SBSR), 2013, Foz do Iguaçu. *Anais...* São José dos Campos: INPE, p. 1885-1892., 2013.

SOUSA, G. M.; SANTOS, F. V.; FERNANDES, M. C. Mapeamento da cobertura da terra da APA Petrópolis/RJ utilizando análise baseada em objeto no sistema InterIMAGE. In.: *Revista Brasileira de Cartografia – RBC*, nº 64/4, p. 517-530, 2012

SOUSA, G.M. *Modelagem do conhecimento aplicada ao estudo da susceptibilidade à ocorrência de incêndios no Parque Nacional do Itatiaia*. Tese de Doutorado. Programa de Pós Graduação em Geografia – PPGG/UFRRJ. 161 f., 2013.

TOMZHINSKI, G.W. (2012) *Análise Geoecológica dos Incêndios Florestais no Parque Nacional do Itatiaia*. Dissertação de Mestrado. Programa de Pós Graduação em Geografia – PPGG/UFRRJ. 13

Trabalho enviado em outubro de 2015.
Trabalho aceito em dezembro de 2015.

Leonardo Herms Maia, Gustavo Mota de Sousa e Manoel do Couto Fernandes,
Uso de mineração de dados para mapeamento da cobertura da terra em imagem Landsat 8 no Parque Nacional do Itatiaia