

## **Classificação da cobertura do solo urbano na área costeira de São Luís – MA, utilizando imagens do satélite *Worldview-II* e técnicas de mineração de dados.**

Ulisses Denache Vieira Souza<sup>1</sup>  
Hermann Johann Heinrich Kux<sup>1</sup>

<sup>1</sup> Instituto Nacional de Pesquisas Espaciais - INPE  
Caixa Postal 515 - 12227-010 - São José dos Campos - SP, Brasil  
{ulisses, hermann}@dsr.inpe.br

**Abstract.** Urban areas are characteristic spaces under dynamic changes, with problems related to planning land cover. Such areas are especially fragile when they are located in coastal regions with mangrove vegetation and dune ecosystems. Remote sensing information and data mining techniques allow the automatic extraction of important information for planning and urban management issues of such areas. Data processing of the new high resolution remote sensing satellite systems considers the GEOBIA paradigm. In this study data from the WorldView-2 satellite. The objective of this study was to evaluate the capability of these datasets for the classification of land cover in urban coastal areas. Two test-sites were considered at the northern section of São Luís city (Maranhão State, Brazil). Initially, tests were made with a classification model, considering only those tools implemented at the InterIMAGE classification software.

**Palavras-chave:** high resolution, Worldview-2, mangrove, alta resolução, Worldview 2, mangue.

### **1. Introdução**

A partir do início da década passada, com a disponibilidade de novos sistemas sensores de alta resolução e a utilização dos Sistemas de Informações Geográficas (SIG's), tornou-se possível a obtenção de imagens que contém alto nível de detalhamento, expandindo assim as possibilidades para o uso de produtos orbitais, devido principalmente à significativa melhoria nas resoluções espaciais e espectrais, semelhantes a fotografias aéreas adquiridas por sensores aerotransportados em baixa altitude (Ehlers, 2007). Com a disponibilidade de novos sistemas sensores de alta resolução e SIG no início desta década, tornou-se possível a obtenção de imagens com grande detalhe.

As informações obtidas por tais sensores permitem estudos e análises de grande interesse para o planejamento bem como para o monitoramento ambiental. Além disso, dados de imagens de sensoriamento remoto apresentam melhor relação custo-benefício em comparação com as fotografias aéreas, principalmente devido ao aumento da disponibilidade de dados de sensoriamento remoto por satélite, associado à elevada taxa de revisita de tais sistemas, e devido à melhoria nas resoluções espaciais, espectrais e radiométricas dos dados de sensoriamento remoto orbital Jensen (2009).

Estudos sobre a cobertura do solo se tornam necessários para o correto gerenciamento dos problemas urbanos. Por cobertura da terra (*land cover*), entende-se como sendo aspectos físicos da superfície de uma dada parcela do solo, por exemplo, grama, concreto, água, etc .

Foram realizados estudos em áreas urbanas para a discriminação do uso e cobertura do solo utilizando imagens de alta resolução, destacando-se Ribeiro (2010), Novack (2009), e Pinho (2005) que desenvolveram metodologias usando o paradigma da classificação orientada a objeto.

A área de estudo está inserida na Ilha do Maranhão, situada ao norte do Estado do Maranhão, região Nordeste do Brasil, no município de São Luís, caracterizando-se por apresentar uma diversidade de habitats costeiros como: mangues, falésias, praias e apicuns que necessitam de um monitoramento freqüente por serem áreas de proteção ambiental sensíveis a alterações devido à ocupação humana.

Este trabalho apresenta um modelo de conhecimento para o mapeamento da cobertura no contexto de uma área urbana costeira (São Luís, Maranhão), classificando imagens do satélite *WorldView-II* utilizando o paradigma de análise de imagens baseada em objeto e técnicas de mineração de dados (*Data Mining*).

## 2. O Satélite WorldView-II

Lançado em Outubro de 2009 o sensor orbital *Worldview-2* possui resolução espacial de 0,46 m na banda pancromática e 1,84 metros nas bandas multiespectrais, além de uma resolução radiométrica de 11 bits (Digital Globe, 2010a), apresentando quatro bandas adicionais em relação ao sistema anterior *QuickBird-2* (faixas do azul, verde, vermelho e infravermelho próximo), a saber: *Coastal*, *Yellow*, *Red Edge*, e *Near Infrared 2*. A disponibilidade de quatro novas bandas no *Worldview-2*, localizadas em faixas específicas do espectro eletromagnético aumentam a capacidade de discriminação e análise dos alvos em estudo Figura 1.

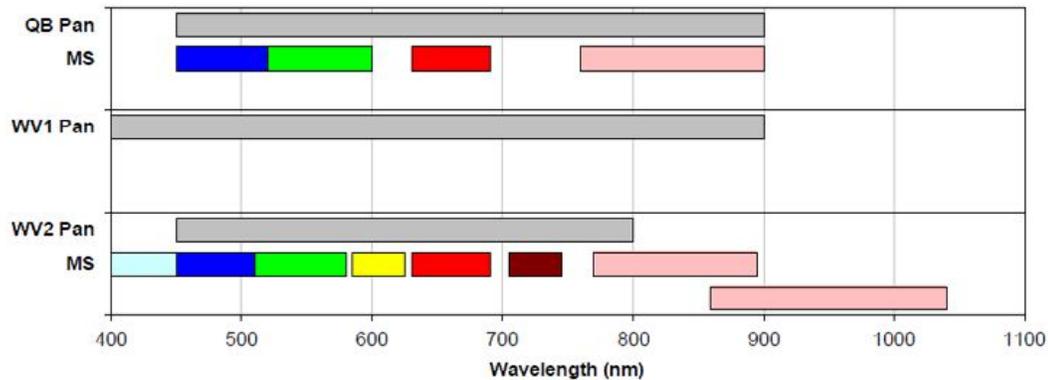


Figura 1. Comparação das bandas dos satélites Quickbird, WorldView-1 e 2  
Fonte: Digital Globe.

Souza et al., (2011) realizaram testes para melhorar a discriminação de tipos vegetais costeiros como manguezais e áreas de apicuns a partir das novas bandas do sensor *WorldView II*. Silva et al., (2011) aplicaram as novas bandas para estudos de batimetria e poluição marinha na área do Rio San Pedro na Espanha. Aplicações utilizando as bandas do *WorldView-II* e propostas de metodologias que incluem desde análise espacial de dados a abordagem baseada em objeto encontram-se em Souza (2012), dentre outros.

## 3. Mineração de Dados

A Mineração de dados é uma aplicação de algoritmos computacionais sobre uma base de dados com o intuito de abstrair conhecimentos novos e úteis. Witten e Frank (2005) relatam que estes algoritmos são fundamentados em técnicas que procuram, segundo determinados paradigmas, explorar os dados de forma a produzir modelos de conhecimento que expressam padrões de comportamento dos dados que podem ser extrapolados para a classificação.

Modelos de conhecimento expressam padrões de comportamento nos dados que podem ser considerados para a classificação ou tomada de decisão em dados diferentes daqueles usados na tarefa de aprendizado, formalizando assim um conhecimento real sobre certo fenômeno Witten e Frank (2005).

Para a escolha de algoritmos de mineração de dados, é fundamental conhecer os tipos de variáveis envolvidas e a inteligibilidade do modelo de conhecimento gerado, para definir a forma de aprendizado do algoritmo, que pode ser: (1) supervisionado, em que o modelo de conhecimento é abstraído a partir de um conjunto de treinamento e avaliado a partir do conjunto de teste e (2) não-supervisionado, quando não existe a informação de saída desejada, e os algoritmos partem dos dados procurando estabelecer relações entre si Goldschmidt e Passos (2005).

#### 4. O InterIMAGE

O InterIMAGE é um software livre para a interpretação de imagens baseado em conhecimento. Desenvolvido a partir da cooperação entre o Instituto Nacional de Pesquisas Espaciais - INPE, o Laboratório de Visão Computacional do Departamento de Engenharia Elétrica da Pontifícia Universidade Católica do Rio de Janeiro (LVC-DEE/PUC-RJ), e a Universidade Leibniz de Hannover (Alemanha). Segundo Costa et al., (2010) e Costa (2009) o InterIMAGE herdou a estrutura de conhecimento, design e os mecanismos de controle do sistema alemão GeoAIDA Buckner al. (2001) este software GeoAIDA consiste num desenvolvimento do AIDA (Automatic Image Data Analyser).

#### 5. Área de Estudo

A área de estudo está inserida na Ilha do Maranhão, situada ao norte do Estado do Maranhão, região Nordeste do Brasil, com uma área total de aproximadamente 831,7 km<sup>2</sup> (e está localizada entre os meridianos W 44°17'55" e W 44°12'18" e os paralelos S 02°30'26" e S 02°27'34". A ilha é composta pelos seguintes municípios: São Luís (capital), São José de Ribamar, Paço do Lumiar e Raposa (Figura 2).



Figura 2. Localização das áreas-testes

Para delimitar as duas áreas-testes estudadas, foram considerados aspectos como diversidade de alvos urbanos na cena, capazes de serem distinguidos pelo sensor, ou seja, objetos sempre presentes em estudos urbanos a partir de dados de sensoriamento remoto, tais como: piscinas, vegetação rasteira e arbórea, diferentes tipos de telhados (diferentes materiais

de cobertura), edifícios com gabaritos de altura diversos, galpões industriais, áreas de solo exposto etc. A presença de tais elementos em uma cena urbana são os motivadores para a construção de uma rede hierárquica e de operadores que contemplem estes diferentes tipos de coberturas Riberiro (2010).

Na área-teste A, estão localizados os conjuntos residenciais de alto padrão horizontais (Calhau e parte da Ponta do Farol) e verticais (área de interface entre os bairros Calhau e Renascença), e também está contida uma porção de praia, um trecho de unidades de conservação, como a Estação Ecológica do Rangedor, faixas de dunas e de restingas (vegetação rasteira) e pequenas áreas comerciais. A Area-Teste A com maior resolução espacial, onde se identifica com maior detalhe a heterogeneidade da cena. Na área-teste B, há ocupações espontâneas (palafitas) na área próxima ao manguezal e ao canal de maré, com a presença também de residências horizontais de baixo padrão, verticais de baixo padrão, áreas em construção de novas vias para a cidade (Avenida Quarto Centenário).

## 6. Metodologia

### 6.1 Material

Este capítulo descreve os dados matriciais e vetoriais, os softwares, aplicativos e equipamentos utilizados e os procedimentos metodológicos executados na realização deste trabalho.

#### 6.1.1 Dados de Entrada

a) Dados Matriciais:

Dois conjuntos de imagens *WorldView-II* do tipo Standard/O2A obtidas no dia 10 de junho de 2010, disponibilizados pela empresa DigitalGlobe. O sensor orbital *Worldview-II* possui resolução espacial de 0,50 m na banda pancromática e 2,00 metros nas bandas multiespectrais.

b) Dados Vetoriais:

- Banco de Dados da cidade de São Luís produzidos durante o projeto “São Luís: Uma leitura da Cidade” Espírito Santo (Org.), (2006) um estudo técnico coordenado pelo Instituto de Pesquisa e Planificação da Cidade, visando subsidiar estudos para aprovação do Plano Diretor da cidade na escala de 1:10.000.
- Curvas de nível com equidistância de 01 metro para a área do município de São Luís disponibilizadas pelo Laboratório de Geoprocessamento da Universidade Federal do Maranhão.
- Limites territoriais do município de São Luís, no formato *ESRI Shapefile* (shp) adaptados da base vetorial disponibilizada pelo IBGE.

#### 6.1.2 Softwares e Aplicativos

- a) Arc GIS 10.0 – Conversão, edição e ajustes topológicos dos dados vetoriais para elaboração do Modelo Digital de Elevação (MDE). Elaboração dos mapas finais;
- b) ENVI 4.7 (ITT, 2009), para mosaicagem, fusão e recorte das áreas testes a partir das imagens do *WorldView-II*, elaboração das matrizes de confusão das classificações feitas no InterIMAGE;
- c) Spectrum Survey 4.22 – utilizado no pós-processamento dos pontos GNSS coletados em campo para a ortorretificação das imagens;

- d) MAPGEO 1.0 – Para tratamento e conversão das coordenadas elipsoidais em ortométricas.
- e) PCI Geomática Ortho Engine 10.3 (PCI GEOMATICS, 2010) - utilizado para a ortorretificação das imagens;
- f) WEKA 3.6.5 – utilizado na mineração de dados (seleção dos atributos e geração das árvores de decisão);
- g) InterIMAGE v1.24 – Para elaboração da rede semântica, definição das regras de pertinência, elaboração dos modelos de conhecimento para classificação de cobertura da terra e uso do solo posterior;
- h) Notepad++ 5.8.7 – Preparação dos dados de mineração através da edição e elaboração dos arquivos ARFF para utilizar no software Weka.

### 6.1.3 Equipamentos

- a) Câmera Fotográfica Sony para registro do local de coleta dos pontos de controle no terreno;
- b) 02 Receptores GNSS SOKKIA numeração 1904 e 1908 – Modelo: Stratus Básico Estático – coleta de pontos em campo para a ortorretificação;
- c) 02 Receptores GPS Garmin – Modelo: 12XL – para localização dos pontos pré-definidos, cálculo da distância de um ponto a outro e do tempo em relação à base.
- d) Trena para cálculo da altura da antena;

## 7. Resultados e Discussões

Neste item são apresentados, avaliados e comparados os resultados das classificações da cobertura dos modelos gerados. Inicialmente, realizou-se uma análise visual dos mapas temáticos em comparação com a imagem original, para verificar se a identificação das classes pela hierarquia proposta é coerente.

O resultado das classificações de cobertura do solo foi avaliado pelo teste estatístico Z quanto à importância de suas diferenças, que foram significativamente melhores do que numa classificação aleatória.

A interpretação visual, a matriz de confusão, o índice *Kappa* e a exatidão global foram utilizados para verificar a qualidade das classificações. Para a distinção entre as classes de cobertura do solo foi utilizado o índice *Kappa* Condicional e exatidões do usuário e produtor Tabela 1

Tabela 1: Índices *Kappa* das classificações.

ID	Área - teste	Exatidão Global	Índice <i>Kappa</i>	Variância - <i>Kappa</i>	Teste Z
01	A - I	0,8467	0,8245	0,000542	5,542
02	A - II	0,8645	0,8354	0,000431	
03	B - I	0,8432	0,8232	0,000324	5,154
03	B - II	0,8678	0,8446	0,000365	

De um modo geral, os testes II (no qual foram inseridos os parâmetros de mineração de dados com intervenção do intérprete) obtiveram melhores valores para o índice *Kappa*-condicional do que os modelos AI e BI (modelo gerado pelo intérprete com base nos atributos extraídos da análise exploratória simples). Segundo Carvalho (2011) as classes “vegetação arbórea” e “vegetação rasteira”, em relação às demais, caracterizam-se por sua fácil distinção, devido aos altos valores de NDVI, apesar da confusão entre ambas. Os gráficos na Figura 3 apresentam os valores obtidos para o índice *Kappa* condicional das classes de cobertura do solo para cada área-teste de estudo (Figura 4 e 5).

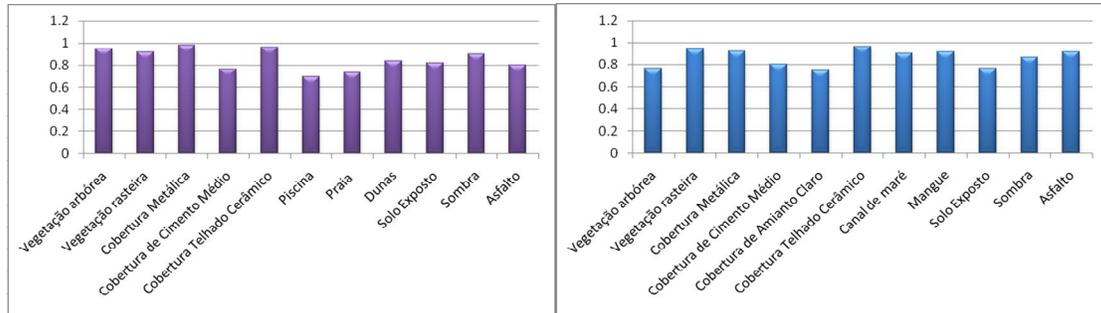


Figura 3. Índices *Kappa* condicional para as classes de cobertura do solo da Área A e B-Teste II.

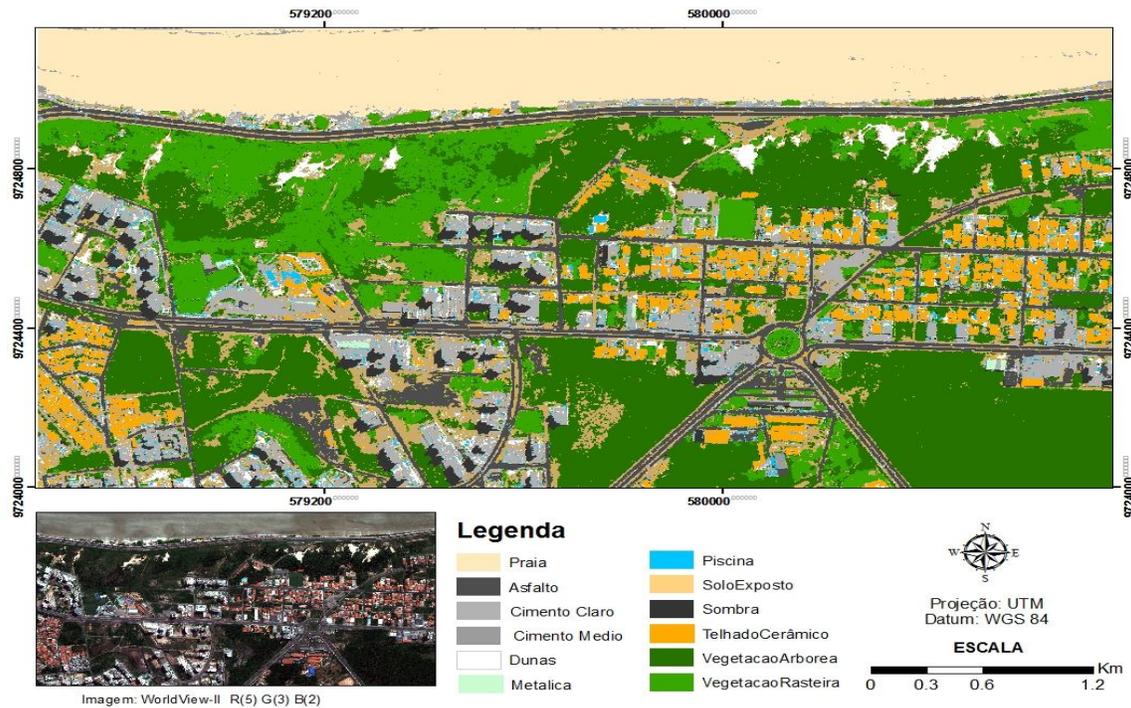


Figura 4. Mapa temático de cobertura do solo: Área A Teste II

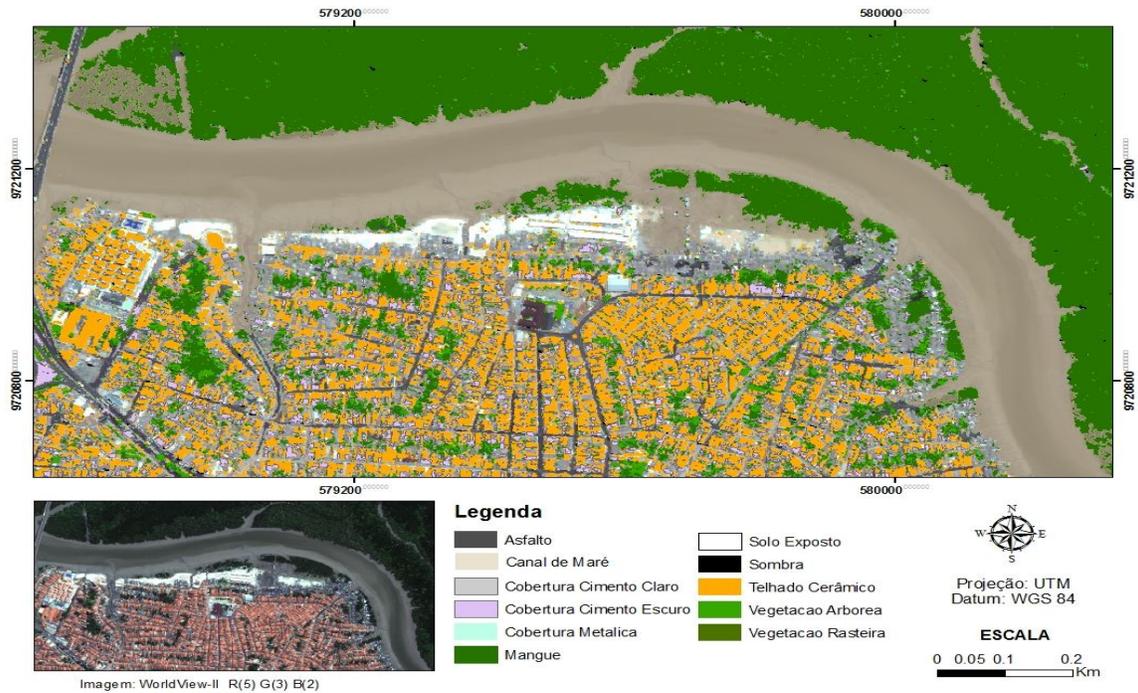


Figura 5. Mapa temático de cobertura do solo: Área B Teste II

## 8. Conclusões

Avaliou-se o uso do modelo de conhecimento com base no mesmo somente pelo especialista, e o modelo com base em técnicas de mineração de dados e intervenção do especialista. Esta hipótese foi testada a partir da avaliação comparativa das classificações das imagens WorldView-II no software InterIMAGE com a implementação dos parâmetros de mineração de dados gerados a partir do software Weka.

A implementação dos modelos I (geração do modelo de classificação por análise exploratória no software InterIMAGE) e II (uso da técnica de mineração de dados por árvores de decisão, buscando um número mínimo de instâncias por folha) indicaram uma maior otimização e potencialidade na ampliação do conhecimento em relação ao modelo da análise exploratória convencional para a classificação de cobertura do solo urbano.

Os melhores resultados para a classificação da cobertura foram obtidos com o modelo AII e BII que utilizaram parâmetros a partir de técnicas de mineração de dados. A exatidão global obtida neste modelo foi de 0,87, e o índice de concordância *Kappa*, de 0,8585. O índice *Kappa* obtido para mapear o uso do solo por quadras das áreas testes foi de 0,8452.

A implementação de uma rede semântica que já considerava uma série de parâmetros obtidos a partir do software para mineração de dados *Weka*, permitiu uma economia de tempo para a análise exploratória dos atributos e limiares, permitindo a sua criação e aplicação neste trabalho.

A técnica de mineração de dados por árvores de decisão facilitou alcançar o principal objetivo deste trabalho, ou seja, a classificação da cobertura e uso do solo urbano, visto que, permitiu obter os principais parâmetros que foram incorporados às regras de decisão do software aplicativo InterIMAGE.

## Agradecimentos

Agradecemos a empresa Digital Globe pela disponibilização das imagens do *Worldview-2* e a Fundação de Amparo a Pesquisa do Maranhão, FAPEMA, pela bolsa de estudos disponibilizada e ao Eng. Paulo Cesar Gurgel de Albuquerque da DSR/INPE pelo importante auxílio na jornada de campo no Maranhão.

## Referências Bibliográficas

- Carvalho, M. V. A. **Classificação da cobertura e uso do solo com imagens Worldview-2 utilizando mineração de dados e análise baseada em objetos**. Dissertação (Mestrado em Sensoriamento Remoto) Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2011, disponível em: <http://mtc-m18.sid.inpe.br/col/sid.inpe.br/mtc-m18@80/2009/08.31.21.23/doc/publicacao.pdf>. Acesso em: 20 de março de 2012
- Digital Globe. **WhitePaper – The benefits of the 8 Spectral Bands of WorldView-II**. Março, 2010a. Disponível em: <[http://Worldview2.digitalglobe.com/docs/Worldview-2\\_8-Band\\_Applications\\_Whitepaper.pdf](http://Worldview2.digitalglobe.com/docs/Worldview-2_8-Band_Applications_Whitepaper.pdf)> Acessado em: 20 de Agosto de 2011.
- Ehlers, M. Sensoriamento Remoto para usuários de SIG – Sistemas Sensores e métodos: entre as exigências do usuário e a realidade. In: Blaschke, T.; Kux, H. (Org.). **Sensoriamento Remoto e SIG avançados – Novos sistemas sensores: métodos inovadores**. São Paulo: Oficina de Textos, 2007. 2ª edição. 303 p. Cap.2, p. 19-38.
- Espírito Santo, M. (Org). **São Luís: Uma leitura da Cidade**. Instituto de Pesquisa e Planificação da Cidade, São Luís: IPPC, 2006. 259 p.
- ITT Visual Information Solutions. **ENVI 4.7, version 4.7**. Boulder, Colorado, EUA, 2009.
- InterIMAGE. Interpreting images freely. Desenvolvido por Laboratório de Visão Computacional da Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, Brasil, 2010. Disponível em: <http://www.lvc.ele.pucRio.br/projects/interimage/> Acesso em 15 jan 2012.
- Jensen, J.R. **Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres**. São José dos Campos: Parêntese, 2009. 604 p.
- Novack, T. Classificação da cobertura da terra e do uso do solo urbano utilizando o sistema InterIMAGE e imagens do sensor Quickbird. 2009. 214p. INPE-16594-TDI/1580. Dissertação (Mestrado em Sensoriamento Remoto). INPE – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2009, disponível em: <http://mtc-m18.sid.inpe.br/col/sid.inpe.br/mtc-m18@80/2009/08.31.21.23/doc/publicacao.pdf>. Acesso em: 20 de outubro de 2011.
- Pinho, C.M.D. Análise orientada a objetos de imagens de satélites de alta resolução espacial aplicada à classificação de cobertura do solo no espaço intraurbano: o caso de São José dos Campos. 2005. 180p. INPE-14183-TDI/1095. Dissertação (Mestrado em Sensoriamento Remoto). INPE – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2005, disponível em: <http://mtc-m17.sid.inpe.br/col/sid.inpe.br/MTC-m13@80/2005/11.23.13.40/doc/publicacao.pdf>. Acesso em: 15 de setembro de 2012.
- Ribeiro, B.M.G. Avaliação de Imagens Worldview-II para o mapeamento da cobertura do solo urbano utilizando o sistema InterIMAGE. 2010. Dissertação (Mestrado em Sensoriamento Remoto). INPE – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2010, disponível em: [http://mtc-m19.sid.inpe.br/col/sid.inpe.br/mtc-m19@80/2010/08.03.12.52.20/doc/Dissertacao\\_BMGR\\_final.pdf](http://mtc-m19.sid.inpe.br/col/sid.inpe.br/mtc-m19@80/2010/08.03.12.52.20/doc/Dissertacao_BMGR_final.pdf). Acesso em: 10 de novembro de 2012.
- Souza, U.D.V. **Classificação da cobertura e do uso do solo urbano de São Luís (MA) com imagens WorldView-2, utilizando mineração de dados e o sistema InterIMAGE**. Dissertação de Mestrado, INPE, São José dos Campos, 2012, 136 p. URL do documento original: <http://urlib.net/8JMKD3MGP7W/3CR8EL8>
- Souza, U.D.V.; Alfaya, F.A.V.S.; Kux, H.J.H. Contribuição das novas bandas espectrais do satélite *WorldView-II* para a Classificação de Tipos Vegetais em Habitats Costeiros: Resultados Iniciais. In (a ser apresentado). XV Simpósio Brasileiro de Sensoriamento Remoto, 01 a 05 de maio 2011, Curitiba, Brasil. **Anais...** Disponível em: <http://marte.dpi.inpe.br/col/dpi.inpe.br/marte/2011/07.20.14.11/doc/p0824.pdf> Acesso em: 03 de abr 2012.
- Weka. Data Mining Software in Java. Desenvolvido pela Universidade de Waikato. Disponível em: <http://www.cs.waikato.ac.nz/ml/weka/> . Acesso em 5 dez 2011.
- Witten, I.H.; Frank, E. **Data mining – practical machine learning tools and techniques** (2.ed.). San Francisco: Morgan Kaufmann, 2005. 558 p. ISBN (0-12-088407-0).